

A Framework for Development of
Data Analysis Protocols
for Ground Water Quality Monitoring

by

Nadine C. Adkins, Ph.D.

June 1993

Technical Report No. 60

Colorado Water Resources Research Institute
Colorado State University
Fort Collins, Colorado 80523
Robert C. Ward, Director

Colorado State University does not discriminate on the basis of race, color, religion, national origin, sex, veteran status or disability, or handicap. The University complies with the Civil Rights Act of 1964, related Executive Orders 11246 and 11375, Title XI of the Rehabilitation Act of 1973, Section 402 of the Vietnam Era Veteran's Readjustment Act of 1974, the Age of Discrimination in Employment Act of 1967, as amended, and all civil rights laws of the State of Colorado. Accordingly, equal opportunity for employment and admission shall be extended to all persons and the University shall promote equal opportunity and treatment through a positive and continuing affirmative action program. The Office of Equal Opportunity is located in Room 21 Spruce Hall. In order to assist Colorado State University in meeting its affirmative action responsibilities, ethnic minorities, women, and other protected class members are encouraged to apply and to so identify themselves.

The contents of this publication are the responsibility of the author(s) alone, and do not necessarily reflect the views of the Institute or Colorado State University.

Copyright © 1993 by Nadine C. Adkins
ALL RIGHTS RESERVED

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher and the author.

Preface

The science of ground water quality monitoring has advanced rapidly in the last decade. Field and laboratory techniques have been developed which allow us to obtain reliable ground water quality data. Data is often collected, however, without a clear view of how it will be used. Ground water quality managers are beginning to realize the value of defining data analysis procedures prior to data collection. This report presents a framework which practitioners can use to develop site specific data analysis protocols. The framework was used to write a data analysis protocol for an IBM semiconductor manufacturing facility in Hopewell Junction, New York.

This report is quite similar to the author's Ph.D. dissertation which was published in 1992 under the direction of Dr. Robert C. Ward. Additional contributors to the dissertation include Dr. Harry F. Bell with IBM, Dr. Hariharan K. Iyer with the Department of Statistics at Colorado State University, and Dr. Jim C. Loftis with the Department of Agricultural and Chemical Engineering at Colorado State University. The author is very grateful for their assistance, encouragement and technical expertise.

Funding for this research was provided by IBM Corporation. IBM's generous contributions allowed us to conduct applied research on statistical analysis of ground water quality data. The project provided an excellent opportunity for cooperation between industry and academia.

This report was typeset by CSWForms of Juneau, Alaska. The author greatly appreciates their patience and attention to detail.

Table of Contents

Abstract	ix
1. Introduction	
Problem Statement	1
Objectives	2
Scope	2
Organization	3
2. Protocols	
Introduction	4
Protocols in the Literature—A Critical Review	4
The Data Analysis Protocol Concept in Water Quality Monitoring	8
The Need for Water Quality Data Analysis Protocols	11
Advantages of Using Water Quality Data Analysis Protocols	11
Characteristics of Effective Water Quality Data Analysis Protocols	12
Conclusions.....	12
3. Identifying Information Goals for Water Quality Monitoring Systems	
Introduction	13
Regulatory Information Goals	13
Monitoring Information Goals.....	18
Statistical Information Goals	19
Conclusions.....	19
4. Handling Attributes of Water Quality Data Records	
Introduction	20
Multiple Observations.....	20
Outliers	20
Changing Sampling Frequencies	20

Missing Values.....	21
Nonnormality	21
Seasonality	24
Censoring	28
Serial Correlation	34
Conclusions.....	40
5. Statistical Analysis of Water Quality Data: Choosing Methods and Analyzing Results	
Introduction	41
Choosing Statistical Analysis Methods.....	41
Graphical Methods	44
Estimation	53
Hypothesis Testing.....	56
Conclusions.....	62
6. A Framework for Development of Ground Water Quality Data Analysis Protocols	
Introduction	63
Components of the Framework.....	63
Summary	65
7. Application of the Design Framework—A Case Study	
Introduction	67
Background.....	67
373 Permitting Process	69
Evaluation of the Framework.....	70
8. Summary and Conclusions	
Summary	72
Conclusions.....	73
Recommendations for Further Work	73
Concluding Remarks.....	74
References.....	75
Data Analysis Protocol for Ground Water Quality Monitoring: IBM East Fishkill Facility.....	colored pages

List of Tables

4-1	Methods for dealing with seasonality (Ward et al., 1988).....	25
4-2	Limits for describing statistical properties of low-level data.....	29
4-3	Definitions of limits for describing statistical properties of low-level data.....	30
4-4	Recommended methods for statistical analysis of censored data. Adapted from Helsel and Hirsch (1992).....	33
5-1	Estimators of distributional properties.....	54
5-2	Estimators of difference in location based on two samples.....	54
5-3	Estimators of linear dependence and monotonic trend.....	55
5-4	Tests for goodness of fit.....	57
5-5	Tests for location—two samples.....	58
5-6	Tests for location—more than two samples.....	58
5-7	Tests for linear dependence and monotonic trend.....	59
5-8	Tests for correlation, equal variance and serial correlation.....	60
5-9	Decisions in hypothesis testing. Adapted from Iman and Conover (1983).....	61

List of Figures

1-1	The role of statistical analysis in relation to modeling and risk assessment.	3
4-1	Graphic depiction of the relationship between the criterion of detection (COD), α and the PDF of H_0	31
4-2	Graphic depiction of the relationship between the criterion of detection (COD), α and β when $COD = LOD$	31
4-3	Graphic depiction of the relationship between the criterion of detection (COD), limit of detection (LOD), α and β when $\alpha = \beta$	31
5-1	Frequency histogram.....	46
5-2	Hanging histogram.....	47
5-3	Stem and leaf diagram.	48
5-4	Boxplot.....	49
5-5	Normal probability plot.....	50
6-1	Main components of the framework for development of data analysis protocols for ground water quality monitoring systems.....	66

Abstract

Protocols for field sampling and laboratory analysis are used on a routine basis to produce reliable ground water quality data. Efforts are now being focused on providing decision makers with the information they need from that data. One method of extracting information from data is statistical analysis. There are, however, no widely accepted protocols for statistically analyzing water quality data. Due to the wide variety of field conditions encountered in ground water monitoring, a general protocol would be of limited use. What is needed is a set of guidelines for writing data analysis protocols which are site specific.

A framework for developing data analysis protocols is presented in this report. The framework is essentially a "how-to" manual for protocol writers. The focus of the framework is analysis of ground water quality data at hazardous waste facilities.

Detailed background information is provided for the framework. Four main issues which are addressed include: information goals, data record attributes, choice of statistical methods, and interpretation of statistical results. There is a great deal of confusion in the water quality community regarding these issues. This report does not attempt to resolve that confusion. Instead, the goal is to sort out areas of conflict and uncertainty, and present them in a clear manner. Recommendations are provided where possible.

The framework was used to write a data analysis protocol for an IBM semiconductor manufacturing plant in Hopewell Junction, New York. The combination of flexibility in the basic framework and the availability of detailed background information was quite effective. It allowed the data analysis protocol to be site specific and scientifically defensible.

* * * * *

Nadine C. Adkins, Ph.D.
Department of Agricultural and Chemical Engineering
Colorado State University
Fort Collins, Colorado 80523
June 1993

PROBLEM STATEMENT

In response to rising environmental concern, the number of ground water quality monitoring systems in the United States is steadily increasing. These monitoring systems are established to meet regulatory as well as internal management needs, but rarely are those needs articulated in terms of the nature of information to be obtained or how the information is to be used in decision making. Consequently, large quantities of data are being collected without a clear definition of program goals, data analysis procedures, reporting formats or types of decisions to be made.

Ground water quality managers are becoming increasingly aware of the need to develop documented strategies (i.e., protocols) for statistically analyzing data obtained from monitoring programs. Data analysis protocols (DAPs) help to ensure that data obtained from monitoring programs can be translated into useful information which meets program goals. Assessing the effectiveness of remediation efforts and detecting the presence of contaminants are examples of program goals.

Ground water quality regulations often contain language similar to the following from the *New York Code of Rules and Regulations*:

In conjunction with a corrective action program, the owner or operator must establish and implement a ground-water monitoring program to demonstrate the effectiveness of the correction action program (Sec. 373-2.6.k.4 of 6NYCRR).

Such regulatory directives usually do not in-

clude details of how to analyze ground water quality data to assess the effectiveness of remediation efforts. It is left to the permittee to decide how to analyze the data. Although a list of proposed statistical methods is generally included in the overall remediation plan submitted by the permittee, it is not standard practice to develop and implement data analysis protocols. As a result, water quality managers often find themselves spending large quantities of money on monitoring programs which do not yield the information they need for decision making.

Ward et al. (1990) have defined water quality monitoring systems in terms of the flow of information through the system. The flow of information begins with sample collection and ends with information utilization as shown below:

1. Sample Collection
2. Laboratory Analysis
3. Data Handling
4. Data Analysis
5. Reporting
6. Information Utilization

The first three components deal primarily with data collection whereas the last three deal with information generation. Researchers have traditionally focused their efforts on data collection issues. Research on data collection has been used to develop standard procedures for sample collection, laboratory analysis and data handling. The use of these standard procedures has resulted in the collection of ground water quality data which has a high degree of

accuracy. Accurate data is, of course, a necessary prerequisite to the generation of useful information.

Water quality researchers and professionals are beginning to switch their focus away from data collection and towards information generation. Some research on information generation activities has already been conducted, particularly in the area of statistical analysis. The research is difficult to apply to monitoring system design, however, because it is so widely scattered.

A few attempts have been made to produce standardized sets of procedures for information generation known as "data analysis protocols." These data analysis protocols (DAPs), however, tend to be incomplete and/or too general to be useful to monitoring system designers.

Due to the wide variety of information needs and site conditions, it is impractical to expect a single DAP to be suitable for all ground water quality monitoring systems. What is really needed is a framework that can be used to develop DAPs which are program specific. No generally acceptable design framework for the development of ground water quality data analysis protocols exists today.

OBJECTIVES

The primary objective of this report is to present a framework for the development of ground water quality data analysis protocols. The protocols are intended to be program specific and should be written during the initial phases of monitoring system design. Application of the data analysis protocols should result in the generation of information which is employed in decision making.

Four main components of the framework which are expanded upon in this report are:

1. Identification of information goals.
2. Handling of data record attributes.
3. Choice of statistical analysis methods.
4. Interpretation of statistical results.

The second objective of this report is to

explore the concept of protocols by summarizing an extensive literature review.

A third objective is to demonstrate the practical application of the DAP design framework by presenting the results of a case study. The framework was used to write a data analysis protocol for an IBM plant in Hopewell Junction, New York.

SCOPE

The intended users of the framework are individuals who will write ground water quality DAPs for hazardous waste facilities. The framework can be used to develop DAPs for internal use and/or for submittal to regulators.

Economic factors are not specifically discussed in this report. It is reasonable to assume, however, that the use of data analysis protocols will produce economic benefits.

The framework should only be used to develop DAPs for chemical ground water quality monitoring programs. Surface water quality and biological ground water quality were not considered in the formulation of the framework.

Although the protocol design framework emphasizes the use of statistical methods of data analysis, there are situations where alternative methods may be more appropriate. Examples of several types of data analysis methods are given in the IBM data analysis protocol which is presented at the back of this report.

This report does not describe the many complexities which contribute to the formulation of water quality laws and regulations. Instead, it emphasizes how regulations should be addressed once they are promulgated. In particular, this report deals with ground water monitoring and data analysis requirements which are placed on industrial facilities.

DAPs developed from the framework should be written prior to data collection. Existing data, however, may be used for data characterization or to confirm analytes.

It is assumed in this report that only high quality data is obtained from the laboratory. The

production of high quality data has been a priority for several years in the field of water quality monitoring. The use of sampling protocols and laboratory analytical protocols (e.g., ASTM procedures) has become routine. Therefore, it is reasonable to assume that hazardous waste sites which have an effective QA/QC program will produce high quality data.

The focus of this report is on the practical application of statistical methods.

The framework is intended for development of DAPs which produce information on **what** the current state of water quality is. The framework is not designed to produce protocols which address the question of **why** water quality is the way it is.

There is a great deal of confusion and uncertainty regarding the use of statistics in the analysis of water quality data. This report attempts to sort out and explain controversial statistical issues rather than resolve them.

It is assumed in this report that readers possess a basic knowledge of statistics. If readers are not familiar with statistics, however, they should still be able understand the report by

referring to an introductory text and/or references noted to contain explanatory detail.

This report addresses the question of how to use statistical analysis to obtain information from data. It does not address the use of modeling or risk assessment. Figure 1-1 shows the role of statistical analysis in relation to modeling and risk assessment.

ORGANIZATION

Protocols are discussed in Chapter 2. Chapters 3 through 5 cover the main aspects of the DAP framework: identification of information goals, handling of data record attributes, choice of statistical analysis methods, and interpretation of results. The DAP framework is presented in Chapter 6 and the case study is discussed in Chapter 7. Conclusions and recommendations for further work are given in the final chapter. The IBM data analysis protocol, which was developed for the case study, is presented on colored paper at the back of the report.

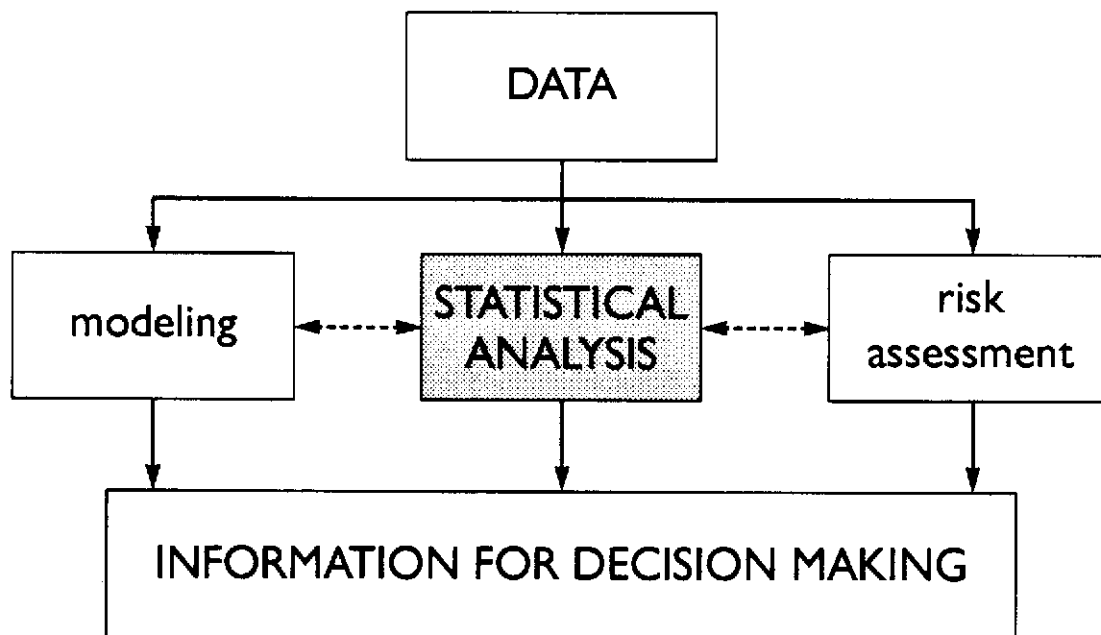


Figure 1-1. The role of statistical analysis in relation to modeling and risk assessment.

INTRODUCTION

The term "protocol" is used in several ways. Many people think of a protocol in terms of diplomacy, either as a signed agreement between negotiating parties or as the proper rules of etiquette used by high officials. The accepted definition in scientific literature, though, is simply: **a standardized set of procedures for doing something**. For example, in hydrology, there are protocols for constructing unit hydrographs and for designing culverts. The American Society for Testing and Materials (ASTM) publishes hundreds of protocols for laboratory and field testing of various substances including water and soil.

An evaluation of what can be learned from existing protocols is presented in the first part of this chapter, followed by a review of the data analysis protocol (DAP) concept in water quality monitoring. Then, the need for ground water quality DAPs is discussed. Finally, the advantages and characteristics of successful ground water quality DAPs are identified.

PROTOCOLS IN THE LITERATURE— A CRITICAL REVIEW

Objective

A literature review of protocols from a variety of disciplines was conducted in order to gain a better understanding of how to develop ground water quality data analysis protocols. The review was not intended to be an exhaus-

tive or precise study, but rather a means to get a general feel for what makes a protocol successful.

Research Strategy

References for the majority of articles included in this literature review were obtained from CDROM catalogs in the CSU research library (using "protocol" or "protocol development" as key words). References were acquired for a large variety of disciplines including microbiology, medicine, toxicology, zoology, environmental engineering, chemistry and anthropology.

A few documents were reviewed which presented protocols but did not identify them as such. *The Consumer Price Index: History and Techniques* (U.S. Dept. of Labor, 1967) and *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities* (U.S. EPA, 1989) are examples.

Results

The literature review yielded a surprisingly large amount of information which can be used to develop ground water quality data analysis protocols. Many disciplines, such as medicine and biology, have been in existence longer than ground water engineering, and are further along in the standardization processes which lead to formulation of protocols.

Observations from the literature review are presented below in a question and answer format. The applicability of these observations to

the development and implementation of ground water quality DAPs is discussed.

Q: *What can one do to ensure the success of a protocol?*

A: *1. Gain an understanding of the problem* - In an article on bioassay protocols (Davis, 1977), the author states "The obvious precursor to standardization is a thorough knowledge of the various chemical, physical and biological factors that affect bioassay results..." This concept is applicable to writing a ground water quality DAP because statisticians are very adamant that the problem must be understood before statistical analysis is conducted (Zahn and Isenberg, 1983).

One approach to gaining understanding of a problem is to review the literature. Literature reviews were conducted for a nutritional risk screening protocol (Hedberg et al., 1988), a hazardous waste reactivity testing protocol (Wolbach et al., 1984), and a protocol for measuring hydrolysis rate constants in aqueous solutions (Ellington et al., 1988).

Implementing a pilot study is another way of acquiring a better understanding of the problem before the final protocol is written (Duke and Merrill, 1981 and Schroder and Taylor, 1980). In water quality monitoring, pilot studies are employed for data characterization, which in turn is used for choosing appropriate statistical methods.

2. Work with others - The chances for success of a protocol are increased if several people are involved in its development. Two methods of involving others are to conduct a workshop and to circulate a draft of the protocol.

Workshops are a commonly used method for getting people together from a variety of disciplines to discuss complex issues. Workshops have been held to discuss the following:

- Protocol standardization for Puget Sound, Washington (Armstrong and Becker, 1986).
- Development of a scientific protocol for

ocean dumpsite designation (Reed and Bierman, 1983).

- Protocol development for potable water reuse criteria (Cotruvo et al., 1982).

Such an approach is appropriate for writing protocols which will be used on a regional or national basis.

A few authors mentioned that they circulated a draft for comments before writing the final protocol (Armstrong and Becker, 1986; Reed and Bierman, 1983; and Schuk, 1986). This is a valuable approach to the development of just about any type of protocol. The chances of the protocol being accepted are increased if users and managers are involved in the development process. Also, extensive feedback minimizes the possibility of omissions or errors in the finished product. Finally, a protocol is by nature a consensus on how something should be done.

Q: *How focused are successful protocols?*

A: *Most protocols encountered in the literature review are narrowly focused* - Some examples are listed here:

- A hospital radiation emergency protocol for Victoria General Hospital (Aldrich, 1988).
- A pronghorn hand-rearing protocol for the Los Angeles Zoo (Brinkley, 1987).
- An analytical protocol for the multimedia characterization of polychlorinated dibenzodioxins and dibenzofurans by high-resolution gas chromatography/high-resolution mass spectrometry (Tondeur et al., 1989).

Because these protocols are narrowly focused, it is possible for the authors to be quite specific. A broadly focused protocol such as one for aquifer decision making (Canter and Knox, 1986) is necessarily general. In fact, the author states "The procedure is not intended to be a set of explicit instructions, but rather a general approach which, when modified, could be applied to a wide variety of ground water pollution problems."

A ground water quality DAP would be the most effective if it were developed for a specif-

ic monitoring program at a specific site. The DAP writer could then take into account the unique hydrogeologic conditions and information goals.

Q: What features enhance the clarity and usefulness of a protocol?

A: 1. Flow charts - Flow charts are often used to visually convey the structure of all or part of a protocol (Reed and Bierman, 1983; Thomas, 1987; and U.S. EPA, 1989). A reader can look at a well designed flowchart and quickly comprehend the basic organization of a process or procedure. An example of simple yet informative flow charts can be seen in a protocol which was developed for ocean disposal site designation (Reed and Bierman, 1983).

2. Glossaries - An EPA handbook titled *Protocols for Short Term Toxicity Screening of Hazardous Waste Sites* (Greene et al., 1988) begins with a glossary. There is enough confusion over statistical and water quality terminology that a glossary would surely be a welcome addition to a ground water quality DAP.

3. Sample forms - An excellent example of the inclusion of sample forms in a protocol is the EPA document, *Interim Protocol for Measuring Hydrolysis Rate Constants in Aqueous Solutions* (Ellington et al., 1988). The forms, which are presented as appendices, are for generation of data and for calculation of rate constants and activation energies. Inclusion of sample forms appears to be a simple and effective approach to ensure that protocol users generate data and perform calculations in a consistent manner. A similar approach may enhance the effectiveness of ground water quality DAPs.

Q: What are some recommendations for development of protocols?

A: 1. Clearly define program goals - The following statement is made in an article which presents a framework for the development of protocols to monitor energy systems in buildings: "Perhaps the most critical activity to the success

of a monitoring project is correctly and accurately stating its goals, objectives, and the specific research questions to be answered" (Misuriello, 1987). There is widespread agreement among water quality professionals that it is important to clearly define the goals of ground water monitoring programs. Statisticians also agree on the importance of well defined goals (Hunter, 1981). It would seem logical, therefore, that a protocol for the statistical analysis of ground water quality data should include a carefully thought out statement of program goals.

2. Use accepted methods - Specification of accepted (i.e., proven and widely used) methods in protocols is promoted in several of the articles which were reviewed. For example, an article which presents a protocol for the identification of toxic fractions in industrial wastewater influents, includes the following statement: "To ensure practicality of the protocol as a routine procedure, the fractionation was based on simple and widely used laboratory techniques..." (Gasith et al., 1988). A similar statement is made in an article on the development of bioassay protocols: "Finally, each test incorporated into the protocol was to be acceptable by the scientific community as valid, reliable and accurate with a sufficiently large existing data base to facilitate the interpretation of results."

Accepted statistical procedures should be specified in a ground water quality DAP to promote routine application of the protocol and to facilitate the interpretation of results. It should be remembered, however, that the use of statistics in analyzing water quality data is a rapidly advancing field. Methods which are "accepted" one year may be replaced by better methods the next year.

3. Allow for revision - A large number of articles which were reviewed included statements indicating that the protocols should be revised as further knowledge is gained. For example, in reference to a protocol for testing microbiological water purifiers: "[it] is intended to be a living document, subject to revision and update as

new knowledge and technology arise" (Schaub and Gerba, 1988). When discussing a protocol for determining lake acidification pathways, the author states "...it provides a general framework which can be challenged and expanded upon through the addition of other evidence..." (Marmorek et al. 1989). A protocol for biological testing was revised and refined to solve problems revealed by pilot studies and implementation (Duke and Merrill, 1981). In reference to a pronghorn hand-rearing protocol: "Over the years we have made modifications to the original protocol as our needs and experience demanded" (Brinkley, 1987).

Allowing for revision of ground water quality DAPs is advisable. As mentioned previously, the use of statistics to analyze ground water quality data is a rapidly evolving field. Also, as more data is obtained in a monitoring program, it may become evident that a change in statistical methodology is appropriate. A systematic revision mechanism should be incorporated directly into the protocol to prevent haphazard and unauthorized changes.

4. Acknowledge limitations - Some of the articles which were reviewed acknowledged limitations of the protocols. For example, in reference to a protocol for determining lake acidification pathways: "The protocol cannot, however, distinguish between low pH lakes which originally contained little DOC, and those which lost it" and "The decision protocol presented in this paper is still very preliminary" (Marmorek et al., 1989). The author of a nutritional risk screening protocol states, "Although the screening program has proved to be an effective tool to identify patients at nutritional risk, it does not substitute for the dietician's clinical subjective experience" (Hedberg, et al., 1989).

It is important to acknowledge the limitations of a ground water quality DAP and the associated statistical procedures so that managers can take them into consideration when making decisions. People tend to place more confidence in statistical results than is warranted.

Q: What are some advantages of using protocols?

A: 1. Protocols save money - Many protocols are designed to save money. For example, when faced with budget cuts, the EPA developed a protocol to more efficiently designate sites for ocean waste disposal (Reed and Bierman, 1983).

A ground water quality DAP can save money in several ways. For one thing, data is not collected unless it is needed for a specific purpose. Also, the maximum amount of information is obtained from the data which is collected.

2. Protocols save time - The need to conduct a project in a timely manner is a common motivator for protocol development. Scientists at the U.S. Department of Agriculture Eastern Research Center wrote a protocol for the accumulation of fatty acid data from multiple tissue samples. The protocol was developed to "allow for an unbiased handling of the samples, proper record-keeping, and—most importantly—the timely completion of an otherwise unmanageable task" (Maxwell and Marmer, 1983).

A protocol for analyzing ground water quality data saves time because unnecessary data is not collected. Furthermore, time is not wasted "reinventing the wheel" whenever data needs to be analyzed.

3. Protocols produce comparable results - A frequently cited advantage of protocols is that data collected by different investigators can be compared. At the start of a regional effort to protect and manage the environment of Puget Sound, it became apparent that data comparability would be a real problem because so many different approaches were being used to sample and analyze the same parameter (Becker and Armstrong, 1988). A series of protocols were written to alleviate this problem. As stated by the protocol authors, "Perhaps the best way to ensure that data collected during different studies are comparable is to encourage all investigators to use standardized sampling and analysis protocols whenever possible."

The advantage of comparable results is perhaps the most evident when monitoring is conducted by many different people on a regional basis. Comparable results are still an advantage, however, even when a ground water quality DAP is written for a specific monitoring program at a single site. Because a protocol documents how statistical analysis results are obtained, future investigators are able to determine if results can be compared.

4. Reliable information is obtained - Several protocols are commended by their authors because they produce reliable information. Some examples are:

- A protocol for measuring hydrolysis rate constants in aqueous solutions (Ellington et al., 1988).
- A protocol for measuring microbial transformation rate constants for suspended bacterial populations in aquatic systems (Steen, 1988).
- A protocol for collection of saliva samples under field conditions (Lipson and Ellison, 1987).

The chances of reliable information being obtained from a ground water quality monitoring program are increased if a DAP is used. The process of writing a DAP forces the author to carefully consider which statistical methods to choose and how to interpret the results. This is in contrast to the haphazard, last-minute application of statistics which often occurs in water quality monitoring. DAPs also lead to more reliable information because they allow data analysis procedures to be audited.

5. Results are reproducible - A few authors cited reproducibility of results as an advantage of their protocol. For example, a treadmill protocol for measurement of aerobic parameters (Cowell, 1989) and a protocol for testing effects of toxic substances on plants (Thompson et al., 1981), both generate reproducible results.

The generation of reproducible results is a significant advantage of using ground water quality DAPs. Too often statistical conclusions

are reported in such a way that the reader has no idea of how they were obtained. The validity of results which are not reproducible is highly questionable.

Summary

Concepts from protocols in other disciplines can be used to develop protocols for analyzing ground water quality data. These concepts are summarized here.

Prior to writing a final protocol, it is essential to understand the issue which is being addressed. Conducting a literature review or a pilot study are two ways to improve understanding.

Working with other people generally contributes to the acceptance and quality of the finished protocol. Input from others can be obtained by holding a workshop and/or circulating a draft.

The majority of successful protocols presented in the literature are narrowly focused.

The clarity and usefulness of protocols can be enhanced with flow charts, glossaries and sample forms.

Recommendations for developing a protocol include:

- Clearly define program goals.
- Use accepted methods.
- Allow for revision.
- Acknowledge limitations.

Well designed protocols save time and money. Also, they produce comparable and reproducible results, as well as reliable information.

THE DATA ANALYSIS PROTOCOL CONCEPT IN WATER QUALITY MONITORING

Protocols which deal with monitoring system design, sampling, and laboratory analysis are routinely used for water quality monitoring. Protocols to statistically analyze water quality data, however, are still in the early stages of development.

Five documents which attempt to standardize various facets of water quality data analysis are reviewed here. The purpose of the review is to gain knowledge which may be helpful in developing DAPs for ground water quality monitoring programs at hazardous waste sites. The five documents are:

1. Nonparametric Tests for Trend Detection in Water Quality Time Series (Berryman et al., 1988).
2. Methodology to Derive Water-Quality Trends for Use by the National Water Summary Program of the U.S. Geological Survey (Lanfear and Alexander, 1990).
3. Water-Quality Data Analysis Protocol Development (Harcum, 1990).
4. Groundwater Quality: A Data Analysis Protocol (Ward et al., 1988).
5. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance (U.S. EPA, 1989).

Nonparametric Tests for Trend Detection in Water Quality Time Series (Berryman et al., 1988). The main objective of this paper is to present a method for systematically choosing the appropriate nonparametric statistical test for analyzing a particular time series. Selection of tests is based on the dependence and sources of dependence in the time series, and on whether the trend is monotonic or step.

The protocol for selecting tests for monotonic trends is presented in both narrative and flowchart form. It can be modified for use with step trends.

The authors stress that the protocol is a general guide which should not be "followed blindly" because user judgment is sometimes required. They point out that the role of user judgment will be diminished in the future as increased knowledge allows for better definition of test selection criteria.

This paper is an excellent reference for anyone who needs to write a DAP which includes time series analysis. The paper's emphasis, however, is strictly on data analysis. It does not address other critical aspects of DAPs such

as identification of information goals or reporting procedures.

Methodology to Derive Water-Quality Trends for Use by the National Water Summary Program of the U.S. Geological Survey (Lanfear and Alexander, 1990). The USGS developed computer software to do the following:

1. Objectively determine the suitability of a water quality record for trend testing.
2. Decide whether to do a monthly, bimonthly or quarterly trend test.
3. Prepare data for analysis.
4. Perform the Seasonal Kendall monotonic trend test and calculate the trend slope.
5. Report the results.

The software (i.e., protocol) was evaluated using data from four states. It was refined until it successfully dealt with all of the data records. After further testing and refinement, the protocol was used to conduct over 50,000 trend tests on data collected from almost 3,000 stations nationwide.

Lanfear and Alexander make this statement regarding the protocol: "Perhaps the most important lesson to be learned from the experiences of developing an automatic trend test is that software must be very 'smart' if it is to cope with the myriad ways in which water-quality data are collected and recorded." Anyone who has dealt with statistical analysis of water quality data would probably agree with that observation.

Water-Quality Data Analysis Protocol Development (Harcum, 1990). In his dissertation, Harcum suggested that a water quality data analysis protocol should address five issues:

1. Identification of information goals and transformation into water quality conditions.
2. Data handling.
3. Identification of data record attributes.
4. Water quality evaluation.
5. Information reporting.

Harcum reviewed the literature and identified those data analysis procedures which have gained wide acceptance. He proposed that a data analysis protocol could be developed simply by combining those procedures. For example, most authors agree that the Mann-Kendall tau test (or variations) should be used for monotonic trend detection, and that the Sen slope estimator should be used for monotonic trend magnitude estimation.

Harcum also proposed that a DAP could be developed by conducting simulation studies. Although he did not actually produce a protocol, Harcum conducted simulation studies which yielded important information which could be used in future protocol development efforts.

From an economic standpoint, it is unrealistic to expect monitoring system designers to conduct simulation studies for DAP development. It is more practical for them to write protocols by combining procedures which have gained wide acceptance.

Groundwater Quality: A Data Analysis Protocol (Ward et al., 1988). A data analysis protocol is presented which contains data preparation procedures, graphical evaluation techniques, and recommended statistical methods. The protocol was developed for analysis of ground water quality data at regulated industrial facilities, but can be modified to analyze other environmental data.

The protocol has several positive features. For one thing, it is easy to use and understand. Also, it uses flowcharts and demonstrates the application of the protocol with a case study.

A disadvantage of the protocol is that it is aimed at industrial facilities in general, and is therefore too broadly focused for direct use in a specific monitoring program. Also, it does not explain the logic behind many of the recommendations which are given.

Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance (U.S. EPA, 1989). This EPA document provides guidance on the selection,

use and interpretation of statistical methods to evaluate ground water quality monitoring data at RCRA (Resource Conservation and Recovery Act) facilities. The manual is a clarification and expansion of amendments made to the RCRA regulations on October 11, 1988. Prior to those amendments, the regulations required that ground water data be analyzed by the Cochran's Approximation to the Behrens-Fisher Student's t-test (CABF). The 1988 amendments eliminated the CABF procedure in favor of five different statistical methods. It is these five methods which are the main topic of discussion in the statistical guidance document.

In many ways, this manual represents the state-of-the-art in ground water quality DAPs. It is the first (and only) widely accepted national effort to standardize the use of statistics in ground water quality monitoring. Several attractive features of the document are listed here:

- The scope, purpose and intended users of the protocol are clearly defined.
- Flowcharts and example calculations are used.
- Background information is provided, such as an overview of the regulations.
- Limitations of the protocol are stated.
- A glossary of statistical terms is given.
- Frequent references for further information are provided.
- The importance of considering environmental factors such as hydrogeology and geochemistry in conjunction with statistical results is emphasized.
- The importance of understanding statistical methods prior to using them is promoted.

Even though the EPA statistical guidance document represents a major advancement over previous standardization efforts, there is still room for improvement. The document would be more practical if it could be used either as a program-specific protocol or as a framework for developing such protocols. It is too flexible and broadly focused to be used directly for a specific monitoring program, and it is not designed to be used as a framework for

protocol development. Although the document emphasizes that site-specific factors must be taken into consideration when choosing, applying and interpreting statistical methods, it does not provide guidance on how to write a protocol which incorporates those factors.

There are also some technical difficulties with the manual. It recommends parametric over nonparametric methods which is contradictory to much of the current literature. (See discussion on nonparametric methods in Chapter 5.) Furthermore, it puts too much emphasis on hypothesis testing at the expense of estimation procedures, and it does not adequately address the concept of statistical power (McBride, et al., 1992).

THE NEED FOR WATER QUALITY DATA ANALYSIS PROTOCOLS

In the past, most research efforts in the area of water quality monitoring have been directed towards data collection rather than towards information generation. Attitudes are changing, however, and people are beginning to realize the importance of obtaining information for decision-making. This attitude is reflected in a statement made by Schubel (1987) regarding estuarine environmental monitoring, "We should spend as much on analyzing data, converting them into information and putting them in the hands of decision makers as we spend on collecting them."

Many authors have recognized the need for standardized procedures to ensure that data obtained from monitoring programs can be translated into useful information which meets program goals. For example, the following statement was made in a RCRA implementation study (U.S. EPA, 1990):

We must define what the RCRA program wants to measure by developing environmentally based goals and objectives. Using these milestones, we must develop an information management plan that serves as the blueprint for collecting information and for developing necessary systems... We must assemble and analyze the data so that we

know where the greatest environmental risks from hazardous waste occur, and can measure the program's success in terms of risks reduced or avoided, rather than the number of activities undertaken (permits issued, inspections performed, etc.).

Based on the above statement, it appears that the use of data analysis protocols on RCRA projects could be quite beneficial. A DAP contains "environmentally based goals and objectives," serves as an "information management plan," and provides a means to "assemble and analyze data" to provide useful insights on the environment.

ADVANTAGES OF USING WATER QUALITY DATA ANALYSIS PROTOCOLS

The use of water quality data analysis protocols can provide the following benefits:

1. Communication between monitoring system designers, environmental managers, and regulators is improved.
2. Continuity of data analysis in the face of employee turnover is facilitated.
3. Existing knowledge about water quality variables is incorporated into the monitoring program.
4. Scientific understanding plays a more important role in decision making.
5. Monitoring system designers, environmental managers, and regulators develop a better understanding of the role of statistics in obtaining water quality information.
6. The subjectivity of statistical analysis is attenuated because statistical methods are specified prior to data collection.
7. Those involved in data collection develop an understanding of the importance of producing high quality data. For example, a DAP may explain why data should not be censored in the laboratory.
8. Economic benefits are realized because only data which contributes to useful information is collected. Also, the maximum amount of information is obtained from the data which is collected.

9. Because they are well documented, data analysis procedures can be reviewed by many knowledgeable people. Too often these procedures exist only in someone's head.
10. The monitoring system design is driven by information goals rather than by politics or short term crises.
11. Data characterization is conducted only if it will contribute to the generation of useful information. For example, demonstration of normality is necessary only if the data is to be analyzed by parametric statistical methods.
12. Final decisions made during regulatory negotiations are put in writing.
13. Future investigators will know exactly how statistical results were obtained.
14. Water quality information is extracted from the data as soon as possible. It is often feasible to use a particular statistical method in the initial stages of monitoring and then switch to a more powerful method as additional data are collected. For example, time series plots may be used to "analyze" trends until enough data is available to apply formal trend tests and/or estimation procedures.
15. Reliable information is obtained because the choice of statistical methods and interpretation of results are carefully considered when the protocol is written.
16. Sampling frequencies are chosen based on information goals.

CHARACTERISTICS OF EFFECTIVE WATER QUALITY DATA ANALYSIS PROTOCOLS

Several characteristics of ground water quality data analysis protocols which con-

tribute to their effectiveness are identified below. The characteristics were chosen based on a literature review of protocols, an examination of water quality data analysis protocols, discussions with water quality professionals, and personal judgment.

1. DAPs should be narrowly focused.
2. They should include features which enhance their clarity and usefulness such as sample forms, glossaries and flowcharts.
3. Accepted statistical methods should be used.
4. Protocol limitations should be stated.
5. There should be agreement among users regarding content of the DAPs.
6. DAPs should address the following topics:
 - Identification of information goals.
 - Handling of data record attributes.
 - Graphical presentation of data.
 - Choice of data analysis methods.
 - Interpretation of results.
 - Information reporting.
 - Protocol revision.

CONCLUSIONS

A literature review of protocols from a variety of disciplines was conducted in order to gain a better understanding of how to develop water quality data analysis protocols (DAPs). The review produced a significant amount of useful information which is summarized in this chapter.

The data analysis protocol concept in water quality monitoring is examined by reviewing five documents which attempt to standardize various facets of water quality data analysis. The need for water quality DAPs and advantages of using them are discussed. Finally, several characteristics which contribute to effective water quality data analysis protocols are listed.

Chapter 3

Identifying Information Goals for Water Quality Monitoring Systems

INTRODUCTION

Information goals provide the basis for data analysis protocols. Data attribute handling, choice of statistical analysis methods, interpretation of results, and reporting are all dependent on what we want to know about water quality conditions.

The importance of information goal identification in the design of water quality monitoring systems has been noted by many authors.

... it is simply a waste of money to monitor without a clear relationship between the information to be produced and its use within the management agency's decision making process (Ward et al., 1990).

A clear statement of the program's monitoring goals, objectives and environmental needs, including both narrow and broader long-term needs, is perhaps the most important section of the [environmental monitoring] strategy (U.S. EPA, 1985).

Many monitoring programs are ineffective because they devote too little attention to the formulation of clear goals and objectives... (NRC, 1990).

We must define what the RCRA program wants to measure by developing environmentally based goals and objectives (U.S. EPA, 1990).

Identification of information goals is a three step process. First, regulatory information goals are identified by meeting with regulators and by reviewing regulations. Then, monitoring information goals are established. Finally, if statistical methods are used to achieve the

monitoring goals, specific statistical information goals are developed. Regulatory, monitoring and statistical information goals are all discussed in this chapter.

REGULATORY INFORMATION GOALS

The Role of Conversation

Discussions between monitoring system designers, environmental managers, and regulators are essential to formulation of regulatory information goals. It is difficult for one person to identify information goals by simply reading the regulations. For one thing, environmental regulations are not written in specific enough terms to be used directly as information goals. It is left up to regulators and industry representatives to decide, on a site specific basis, what information is needed from the monitoring system and what the monitoring system can produce.

Secondly, regulations are frequently not based on a true understanding of the problem. This deficiency is reflected in a remark made by a staff assistant in the U.S. House of Representatives: "It is a complete and utter disaster when you begin to look at the data upon which some people on Capitol Hill are talking about basing [ground water quality] regulation—a crap shoot in many cases... We went into this issue thinking we had this area pretty well boxed in and I've personally come out so confused I don't know where to go next" (Nelson and Dowdy, 1988). An understanding

of the problem is necessary for the identification of regulatory information goals, and as stated by Wurman (1989), "Conversations are organic: their very structure is a give-and-take that allows understanding to happen."

Reviewing the Laws and Regulations

Regulations should be thoroughly reviewed prior to defining information goals. It may also be beneficial to review the applicable statutes in order to gain further insight into the intent of the regulations.

Despite the fact that laws and regulations can be vague and unscientific, it is important to review them carefully because they provide the basic "rules of the game." Although there is often plenty of room for negotiation, final water quality decisions must be made within the framework of the regulations.

Statutes are published in the *U.S. Code*, a multi-volume set of books which is divided into 50 titles. Each title covers a broad subject area. Most laws which pertain to ground water monitoring, including RCRA and CERCLA, can be found in Title 42—The Public Health and Welfare.

Federal agency regulations are printed in the *Code of Federal Regulations (C.F.R.)*, which is also a multi-volume set of books divided into 50 titles. EPA regulations are published in Title 40—Protection of the Environment. The *C.F.R.* is updated daily in the *Federal Register*.

Resource Conservation and Recovery Act (RCRA)

General. The Resource Conservation and Recovery Act (RCRA) regulates handling and disposal of solid waste from the point of generation to ultimate disposal (Jorgensen, 1989). RCRA is the first comprehensive piece of federal legislation which addresses the problems of hazardous waste (Hall et al., 1987). Protection of ground water from hazardous waste leachates is covered in detail by RCRA.

The first federal law which required environmentally sound solid waste disposal practices was the Solid Waste Disposal Act

(SWDA) of 1965 (U.S. EPA, 1990). The SWDA was "amended" (virtually rewritten) in 1976 by the Resource Conservation and Recovery Act. The acronym "RCRA" generally refers to the 1976 Act as codified and amended (Cooke et al., 1987a). Major RCRA amendments include the Used Oil Recycling Act of 1980, the Solid Waste Disposal Act Amendments of 1980, and the Hazardous and Solid Waste Amendments (HSWA) of 1984 (Cooke et al., 1987a). The HSWA of 1984 substantially broadened the scope and coverage of RCRA.

Ground Water Quality Information Goals Expressed in the RCRA Statute. The RCRA statute [*U.S. Code*, Title 42, Sections 6901-6991(i)] mandates that EPA promulgate regulations to require ground water quality monitoring systems at a variety of hazardous waste sites.

One of the eleven major objectives of the RCRA statute refers to the preservation of water resources:

The objectives of this chapter are to promote the protection of health and the environment and to conserve valuable material and energy resources by—promoting the demonstration, construction, and application of solid waste management, resource recovery, and resource conservation systems which preserve and enhance the quality of air, water, and land resources.

EPA is authorized to require the owner of a hazardous waste site which "may present a substantial hazard to human health or the environment, (to monitor) to ascertain the nature and extent of such hazard." It is clear that EPA had very little guidance on what the information goals of ground water quality monitoring should be.

Ground Water Quality Information Goals Expressed in the RCRA Regulations. RCRA hazardous waste regulations which apply to treatment, storage and disposal (TSD) facilities will be discussed in this report. The TSD regulations make up the most detailed and complex category of RCRA regulations (Hall et al., 1987).

There are TSD ground water regulations for interim facilities and for permitted facilities. An interim facility is one which has not yet received a permit but has complied with certain generic performance standards allowing it to remain in operation (U.S. EPA, 1990). The owner or operator (referred to collectively as "owner") of an interim status facility is responsible for interpretation and application of ground water regulations, whereas the owner of a permitted facility can refer to site-specific details in the permit (Hall et al., 1987). Ground water regulations and information goals for both categories of TSD facilities are described below.

- **Interim status TSD facilities**

Ground water regulations for interim status TSD facilities are set forth in Title 40 of the *Code of Federal Regulations*, Part 265, Subpart F. There are two stages of monitoring for interim facilities: detection and, if necessary, assessment monitoring.

For detection monitoring, the facility owner is required to monitor for three sets of parameters:

1. Parameters characterizing the suitability of ground water as a drinking water supply.
2. Parameters establishing ground water quality.
3. Parameters used as indicators of ground water contamination (indicator parameters).

If comparisons of indicator parameters between upgradient and downgradient wells show a significant increase (or pH decrease), then the owner must resample and verify the results. If the results are still significant, the owner is required to move into the assessment monitoring phase.

During the assessment monitoring phase, the owner is required, at a minimum, to determine, "(i) The rate and extent of migration of the hazardous waste or hazardous waste constituents in the ground water; and (ii) The concentrations of the hazardous waste or hazardous waste constituents in the ground water." Based on these determinations, the owner must decide if haz-

ardous waste or hazardous waste constituents from the facility have entered the ground water.

The process of identifying ground water quality information goals can begin by reviewing the following statements given in Section 264 Subpart F of the RCRA regulations:

...the owner ... must implement a ground-water monitoring program capable of determining the facility's impact on the quality of ground water in the uppermost aquifer underlying the facility...

[The] number, locations and depths [of downgradient monitoring wells] must ensure that they immediately detect any statistically significant amounts of hazardous waste or hazardous waste constituents that migrate from the waste management area to the uppermost aquifer.

[The owner is required to determine] (i) The rate and extent of migration of the hazardous waste or hazardous waste constituents in the ground water; and (ii) The concentrations of the hazardous waste or hazardous waste constituents in the ground water.

- **Permitted TSD facilities**

Ground water regulations for permitted TSD facilities are set forth in Title 40 of the *Code of Federal Regulations*, Part 264, Subpart F. There are three stages of monitoring for permitted facilities: detection monitoring and, if necessary, compliance and corrective action monitoring.

For detection monitoring the owner is required to monitor for "indicator parameters, waste constituents, or reaction products that provide a reliable indication of the presence of hazardous constituents in ground water." The parameters or constituents to be monitored are specified in the facility permit. If the owner determines that there is statistically significant evidence of contamination for the specified chemical parameters or hazardous constituents, then he or she must "immediately sample the ground water in all monitoring wells and determine whether constituents in the list of Appendix IX of Part 264 are present, and if so, in what concentration." If Appendix IX hazardous constituents are detected (a second analysis is permitted), then they will form the basis for compliance monitoring.

For compliance monitoring, the owner must continue to monitor for Appendix IX constituents and for each chemical parameter or hazardous constituent listed in the permit. In addition, the owner must determine whether regulated units are in compliance with the ground water protection standard specified in the permit. If the ground water protection standard is violated, then the owner must submit a "plan for a ground water monitoring program that will demonstrate the effectiveness of the corrective action."

A corrective action monitoring program must be as effective as the compliance monitoring program in determining compliance with the ground water protection standard. In addition, it must be capable of "determining the success of a corrective action program."

The following statements given in Section 264 Subpart F of the RCRA regulations can be used to begin the process of identifying information goals:

The ground-water monitoring system must ... yield ground-water samples from the uppermost aquifer that: (1) Represent the quality of background water that has not been affected by leakage from a regulated unit;... (2) Represent the quality of ground water passing the point of compliance; (3) Allow for the detection of contamination when constituents have migrated from the waste management area to the uppermost aquifer.

The ground-water monitoring program must ... ensure monitoring results that provide a reliable indication of ground-water quality below the waste management area.

The sample size shall be as large as necessary to ensure with reasonable confidence that a contaminant release to ground water from a facility will be detected.

Use of any of the following statistical methods must be protective of human health and the environment ... (1) A parametric analysis of variance (ANOVA) followed by multiple comparison procedures to identify statistically significant evidence of contamination.

The owner or operator must monitor for indicator parameters (e.g., specific conductance, total organic carbon, or total organic halogen), waste

constituents, or reaction products that provide a reliable indication of the presence of hazardous constituents in ground water.

The owner or operator must determine whether there is statistically significant evidence of increased contamination at each monitoring well at the compliance point ...

... the owner or operator must establish and implement a ground-water monitoring program to demonstrate the effectiveness of the corrective action program.

Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)

General. The Comprehensive Environmental Response, Compensation and Liability Act (CERCLA or Superfund) established a program to identify, investigate, clean up, and impose liability for abandoned hazardous waste sites (Cooke et al., 1987b). CERCLA was passed on December 11, 1980 and was amended by the Superfund Amendments and Reauthorization Act (SARA) on October 17, 1986 (Hall et al., 1990). One of the primary reasons for enactment of Superfund was that RCRA does not authorize EPA to respond to toxic releases at abandoned hazardous waste sites (Glass, 1988). Except for the "imminent hazard" provisions, RCRA does not deal with abandoned facilities (Hall et al., 1990).

The 1980 Superfund statute was a hastily drawn-up, compromise bill characterized by "numerous ambiguities, omissions, and poorly drafted provisions" (Cooke et al., 1987b). A major controversy over the statute became known as the "How Clean is Clean?" question because Congress did not adequately address cleanup standards (Brown, 1990). The 1986 SARA reflects Congressional intent to resolve ambiguities in the 1980 statute. As well as modifying the existing legislation, SARA also added several provisions including a section on cleanup standards (Brown, 1990). The section on cleanup standards, however, is primarily narrative rather than quantitative and leaves EPA considerable discretion to fill in the details.

Ground Water Quality Information Goals Expressed in the CERCLA Statute. The CERCLA statute (as revised by SARA) may be found in the *U.S. Code*, Title 42, Sections 9601-9675. Section 9605 requires EPA to incorporate the provisions of SARA which relate to remedial action, into the National Contingency Plan (NCP). Specifically, CERCLA states that the NCP "...shall establish procedures and standards for responding to releases of hazardous substances, pollutants, and contaminants..."

Ground water quality information goals are referred to in the following statements given in the CERCLA statute:

In assessing alternative remedial actions, the President shall, at a minimum, take into account: ... (B) the goals, objectives, and requirements of the Solid Waste Disposal Act.

Such remedial action shall require a level or standard of control which at least attains Maximum Contaminant Level Goals established under the Safe Drinking Water Act ... and water quality criteria established under section 304 or 303 of the Clean Water Act [where relevant and appropriate].

In determining whether or not any water quality criteria under the Clean Water Act is relevant and appropriate..., the President shall consider the designated or potential use of the surface or groundwater...

As with RCRA, it is evident that EPA had limited guidance on what the information goals of ground water quality monitoring should be.

Ground Water Quality Information Goals Expressed in the CERCLA Regulations. CERCLA implementation policy is codified by EPA in the National Contingency Plan (NCP) (U.S. EPA, 1988). The NCP specifies procedures, criteria and responsibilities for conducting response actions at Superfund sites (Cooke et al., 1987b). The most recent revision of the NCP can be found in Title 40 of the *Code of Federal Regulations*, Part 300.

In general, CERCLA does not contain the detailed type of regulations which are in RCRA. Instead, CERCLA uses broad terminology to outline the organizational structure and

procedures for responding to releases. Procedures for responding to hazardous substance release will be briefly discussed here in terms of the applications of ground water quality monitoring.

Hazardous substance response is addressed in Subpart E of the NCP. There are two types of responses at Superfund sites: removal action or remedial action. Removal action refers to the "the cleanup or removal of released hazardous substances from the environment." Remedial action refers to a permanent remedy which is taken instead of, or in addition to, a removal action. Ground water monitoring generally plays less of a role for removal action than it does for remedial action. If the decision has been made to implement removal action and there will be a planning period of at least six months, then a sampling and analysis plan is required for any environmental monitoring which is conducted. The plan "shall provide a process for obtaining data of sufficient quality and quantity to satisfy data needs."

The NCP outlines several steps in the remedial action process. The steps are as follows:

1. Remedial site evaluation.
2. Remedial investigation.
3. Feasibility study.
4. Remedy selection.
5. Remedial design/remedial action.

The remedial site evaluation involves data collection and evaluation of releases of hazardous substances, pollutants, or contaminants. Although existing data is relied upon heavily at this stage, limited field sampling may also be necessary. A sampling and analysis plan is required if field sampling is conducted.

The purpose of the next step, remedial investigation (RI), is to "determine the nature and extent of the problem presented by the release." Because data collection and site characterization are emphasized in the RI, ground water monitoring often plays a large role.

The main objective of the feasibility study is to "develop and evaluate options for remedial

action." The feasibility study should make use of data gathered in the remedial investigation. Information from the feasibility study is then used to select a remedial action which provides the best balance of trade-offs based on several criteria.

The remedial design results in a "detailed set of plans and specifications for implementation of the remedial action." A ground water monitoring network design may be included in the plans. Once the remedial action is in operation, ground water monitoring may be needed to assess the progress of remediation.

The following statements given in the NCP can be used to begin the process of identifying information goals:

The national goal of the remedy selection process is to select remedies that are protective of human health and the environment, that maintain protection over time, and that minimize untreated waste.

[One of EPA's expectations which should be considered in developing appropriate remedial alternatives is to] return usable ground waters to their beneficial uses wherever practical, within a time frame that is reasonable given the particular circumstances of the site.

The purpose of the remedial investigation {RI} is to collect data necessary to adequately characterize the site for the purpose of developing and evaluating effective remedial alternatives.

... a restoration activity will be considered administratively "complete" when: (i) Measures restore ground- or surface-water quality to a level that assures protection of human health and the environment;...

MONITORING INFORMATION GOALS

Monitoring information goals are qualitative statements which describe specific information expectations of the monitoring program. They provide the underlying framework around which the data analysis protocol is organized.

Monitoring information goals are more specific than regulatory information goals. For

example, a regulatory goal may be: **assess progress of remediation**, and the corresponding monitoring goals may be: **determine the lateral extent of contamination plumes and determine the mass of contaminants removed from each remediation area.**

Monitoring goals do not always have to correspond to a regulatory goal. In some cases, management may need information from the monitoring program which is not required by law. Statistical goals generally evolve from the monitoring goals as the means for obtaining desired information are developed.

Although identification of monitoring information goals is site specific, there are a few underlying principles which are relevant in almost all situations. These are discussed here:

1. **Involve others in the formulation of monitoring information goals.**

During the process of formulating information goals, the DAP writer should request input from others who are involved in the monitoring program. These may include regulators, managers, field samplers, chemists and environmental compliance staff. The chances for a successful protocol are greatly increased if a consensus can be reached regarding the content of monitoring information goals.

2. **Don't include statistical terminology in the statement of monitoring information goals.**

There are two main reasons why statistics should not be included in the statement of monitoring information goals. First, monitoring information goals are intended to be clear, concise statements which can be easily understood and debated by anyone who is involved in the monitoring program. Many people are not comfortable with the language of statistics.

The second and most important reason to exclude statistical terminology is that statistics is a tool for achieving the monitoring information goal. It is not part of the goal

itself. For example, a monitoring goal may be: **determine if concentrations of TCE at Well A are higher than background levels.** Once the monitoring goal is established, statistics can be used to quantify the probability that a given concentration difference for a given sample size is due to chance.

3. Be prepared to revise.

Formulation of information goals is a dynamic process. Ongoing communication between individuals involved with the monitoring program will most likely result in changes to the originally drafted goals. Economics can also be a factor. Once the goals are identified, it may become evident that there is not enough money available to achieve them all. Further revision of monitoring information goals may then be necessary.

STATISTICAL INFORMATION GOALS

Statistical information goals are complete, detailed statements which describe statistical intent. An example is: **detect monotonic, gradual trends in TCE concentrations equal to the standard deviation of the detrended data at the 95 percent significance level over a five year period of quarterly sampling.**

Statistical information goals for hypothesis tests should include three of the following parameters:

- Sample size.
- Significance level.
- Probability that a given difference will be detected (i.e. power).
- Magnitude of the differences to be detected.

Assuming that the population standard deviation can be estimated, the above parameters are related in such a way that any three of them will define the fourth.

Ideally, a data analysis protocol should be written prior to sample collection. If this is done, sampling size can be tailored to meet statistical information goals. The power can be determined for a particular sampling frequen-

cy, and if it is not acceptable, the frequency can be adjusted. If the samples have already been collected, however, the protocol writer has limited control over power.

It is sometimes difficult to formulate a statistical information goal which closely matches the monitoring goal when using standard methods of hypothesis testing. For example, it is conventional for the null hypothesis to reflect the status quo of "no difference." The actual hypothesis of interest, however, may be whether a particular variable falls inside or outside a range of practical importance. Issues related to hypothesis testing are addressed in Chapter 5.

Although the components of a statistical information goal will vary depending on the statistical method which is chosen, there are two principles which are generally applicable. First, the statistical information goal should be specific and complete. Secondly, it should reflect the monitoring goal as closely as possible given the restrictions of the statistical method.

CONCLUSIONS

The development of DAPs should be driven by the information goals. Decisions on how to handle data attributes, which statistical methods to use, and how to interpret and report results should be consistent with information goals.

The process of identifying information goals is highly dependent upon communication between DAP users. It is a process which involves asking questions, sharing ideas, and developing an understanding of the problem.

Of the three types of information goals which were discussed (regulatory, monitoring and statistical), regulatory information goals are the most difficult to identify because the pertinent regulations may be vague and complex. Monitoring information goals, however, are the most important because they form the underlying framework of the data analysis protocol. Statistical information goals should be specific and complete, reflecting the monitoring goals as closely as possible.

Chapter 4

Handling Attributes of Water Quality Data Records

INTRODUCTION

Data record attributes are characteristics of data which can complicate statistical analysis (Bell and Delong, 1988). The following data record attributes are discussed in this chapter:

- multiple observations
- outliers
- changing sampling frequencies
- missing values
- nonnormality
- seasonality
- censoring
- serial correlation

The attributes are generally discussed in order of their complexity. Nonnormality, seasonality, censoring and serial correlation are given the most emphasis.

MULTIPLE OBSERVATIONS

Multiple observations occur when more than one analytical result is recorded for the same time period. This generally happens when replicate samples are collected for QA/QC purposes. QA/QC data should be stored and analyzed with the rest of the data. If a single value is needed for statistical analysis, the multiple values should be averaged.

OUTLIERS

Outliers are values which are obviously higher or lower than the majority of data. They

may be caused by (Helsel and Hirsch, 1992):

- A measurement or recording error.
- An observation from a population not similar to that of most of the data.
- A rare event from a single skewed population.

Outliers which result from either of the last two causes are true observations and should not be discarded.

Erroneous observations may be the result of several factors including sample contamination, failure of laboratory equipment, mistakes by the chemist, or sloppy data entry. Statistical methods can give misleading information when erroneous observations are included in the data analysis. If there is evidence to show that an outlier is an erroneous observation, it should be discarded. Otherwise, it should be retained and used in statistical analysis applications along with the other data (Harcum, 1990). Erroneous observations can be reduced with an aggressive QA/QC program.

CHANGING SAMPLING FREQUENCIES

A variety of factors can cause changing sampling frequencies. Some of these factors are:

- Increased or decreased funding.
- Changing regulatory requirements due to employee turnover in the regulators' offices.
- Modified management priorities resulting from the discovery of new contaminants.

Statistical methods which require equally spaced sampling intervals are not directly applicable to data which has been collected with changing sampling frequencies. The alternatives are to collapse or to exclude data so that all observations are equally spaced. Collapsing data will cause the periods with higher sampling frequencies to have lower variances. Excluding data causes information loss.

Changing sampling frequencies should be prevented because of the difficulties they cause for statistical analysis. Writing a data analysis protocol prior to startup of the monitoring program is a crucial step towards prevention. The DAP will help to ensure that sampling frequencies are chosen based on careful, long-range planning rather than spur-of-the-moment decisions. Also, the DAP can be used to educate managers and regulators of the importance of equally spaced samples.

MISSING VALUES

Missing values can be either random or systematic (Harcum, 1990). Random missing values may occur due to factors such as equipment failure, misplaced samples or test results, inclement weather, employee illness or war. Systematic missing values are often weather dependent. For example, wells may consistently dry up during the summer or be inaccessible during the winter. Changing sampling frequencies also result in systematic missing values (Lettenmaier et al., 1991).

Some statistical analysis techniques which require regularly spaced samples in time or equal sample sizes, cannot be applied to data records which have missing values. Nonparametric methods can usually accommodate random missing data. Systematic missing values, however, can present special problems (Lettenmaier et al., 1991).

Replacing missing values with numerical estimates is not recommended due to possible bias of statistical results. For most situations, the best alternative is to use nonparametric

methods which can accommodate missing data. If there is a large percentage of missing values in a data record or if the missing values are systematic, the best alternative may be to collapse the data prior to statistical analysis.

Harcum (1990) investigated the effect of missing values on trend detection. He addressed the following question: "How many missing values are needed in a data set before it is necessary to collapse the data from monthly to quarterly values?" Harcum made the following conclusions:

- When applying the Mann-Kendall tau or Seasonal Kendall tau tests, collapse monthly data to quarterly values if more than 50 percent of the monthly data are missing .
- When applying the Seasonal Kendall tau test with correction for serial correlation, collapse monthly data to quarterly values if more than 40 percent of the monthly data are missing .
- When there are only five years of record and more than 50 percent missing values, there is not a good alternative.

Harcum's recommendations are the most relevant to someone who is attempting to analyze existing data. A data analysis protocol writer should not have to deal with such large percentages of missing values if the monitoring program is well planned and operated.

Every effort should be made to prevent missing values because of the complications they cause for statistical analysis. An effective QA/QC program is necessary. Also, those involved in field sampling and laboratory analysis should realize the importance of preventing missing values.

NONNORMALITY

Introduction

Water quality data are frequently right skewed and therefore violate the assumption of normality. Most parametric statistical methods which are used to analyze water quality

data assume that the data are normally distributed. The power of such methods is sacrificed if the assumption of normality is violated.

To avoid the problem of loss of power due to nonnormal data, investigators are turning to nonparametric techniques. For some applications, however, nonparametric methods are not available. Also, regulations may require that parametric approaches be used. If parametric approaches are used, the distribution of the data should be investigated.

Distributional Characteristics of Water Quality Data

It is common for water quality data to be nonnormally distributed. They are often right (positive) skewed because they have a lower bound of zero with infrequent high values. The presence of nondetects accentuates right skewness.

Several investigations of the distributional properties of water quality data records have been conducted. Some of these studies are summarized here.

Montgomery et al. (1987) examined 172 ground water quality records for normality.¹ They found that 106 (62 percent) of the records were nonnormally distributed. The nonnormal records tended to be right skewed with the degree of skewness varying widely.

Gilliom and Helsel (1986) examined the statistical characteristics of trace constituent concentrations in samples collected at U.S. Geological Survey river water quality monitoring stations. For 482 uncensored data sets, sample skewness ranged from -0.8 to 5.2 with a median of 1.8. Only 6 percent of the values were negative. A normal distribution is, by definition, symmetric and therefore has a skewness of zero. Gilliom and Helsel's results show that water quality data are frequently right skewed.

Loftis et al. (1989) checked water quality records from lakes for normality.² They found that only about 20 percent of the records displayed significant nonnormality. They repeated their analysis after log transforming the

data and also after removing quarterly means. (Log transformations can sometimes make data more normal, and removing quarterly means can decrease the influence of seasonality thereby increasing the power of tests to detect nonnormality.) The authors concluded that neither action had much effect on the percentage of records which appeared to be nonnormal.

Effects of Nonnormality on Statistical Analysis

Two undesirable effects can occur when significance tests which assume normality are applied to nonnormal data. The first effect is a distortion of Type I error (Snedecor and Cochran, 1989). For example, if the nominal significance level is set at 0.05 and the null hypothesis is true, H_0 should be rejected 5 percent of the time. If the assumption of normality is not satisfied, however, H_0 may be rejected perhaps 3.6 percent or 8.5 percent of the time.

The second effect which can occur when significance tests which assume normality are applied to nonnormal data is loss of power (Conover and Iman, 1976 and Helsel, 1987). Low power is a potentially serious problem because important changes or differences in water quality may go unrecognized.

When tests are said to be robust against nonnormality, it generally means that nonnormality does not cause much distortion of Type I error (Conover and Iman, 1976). For example, because the nominal significance level of

¹ Montgomery et al. (1987) used frequency histograms, normal probability plots, the chi-squared goodness-of-fit test, and the skewness test to examine data records for normality. They used a significance level of 5% for the two hypothesis tests.

² Loftis et al. (1989) used two hypothesis tests to check for the presence of nonnormality. The null and alternative hypotheses for each test were: (1) H_0 : skewness = 0 and H_a : skewness \neq 0, and (2) H_0 : kurtosis = 3.0 and H_a : kurtosis \neq 0. Both tests were applied at significance levels of 10% and 2% (i.e., 5% and 1% for each tail).

the t-test is preserved for large sample sizes, the t-test is often cited as being robust against nonnormality. This statement ignores the loss of power which can occur when the assumption of normality is not satisfied (Helsel, 1987).

Ways to Deal With Nonnormality

Potentially nonnormal data can be dealt with the following ways:

- Use nonparametric statistical methods.
- Test for normality and if the data is normal, use parametric methods.
- Test for normality and if the data is nonnormal, use transformations. Proceed with parametric analysis if the transformed data is normal.
- Do a sensitivity analysis to examine the effects of nonnormality on the test you want to use. If the effects are acceptable, proceed with the test.

The first alternative, using nonparametric methods, has become the most accepted approach to statistically analyzing water quality data. One reason that nonparametric methods are so popular is because they do not assume normality. (Additional advantages are discussed in Chapter 5.) Hollander and Wolfe (1973) made the following statement regarding the efficiency³ of nonparametric methods:

More often than not, the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal (the home court of normal theory methods), and they can be mildly and wildly more efficient than these competitors when the underlying populations are not normal.

Bradley (1968) and Hirsch et al. (1991) have made similar conclusions regarding the efficiency of nonparametric methods.

The normality of data sets should be investigated prior to conducting parametric tests.⁴ If the data is shown to be nonnormal, transformations can be calculated and the data reexamined for normality (Helsel, 1987). If the data is still nonnormal or if transformations are not desirable, a sensitivity analysis can be conduct-

ed to quantify the effects of violating the normality assumption (Hirsch and Slack, 1984).

There are at least two problems associated with tests for normality. One problem is that the tests have very low power for small (<30) sample sizes (Helsel and Hirsch, 1988). This means that only extreme cases of nonnormality will be detected if the sample size is small. The second problem is that even if a test for normality has adequate power, acceptance of the null hypothesis does not prove that the data is normal (Helsel and Hirsch, 1992 and Montgomery and Reckhow, 1984). It merely indicates that there is not enough evidence to indicate nonnormality.

Transformations are generally not desirable for multiple data sets (Hirsch et al., 1991). Each data set should be examined separately to choose the best transformation. This can be quite time consuming. Also, if several transformations are chosen, it can be difficult to compare statistical analysis results between data sets. Another drawback of transformations is that the transformed data must be tested for normality (Helsel, 1987). It is not safe to assume that transformed data is automatically normal and therefore acceptable for analysis by parametric methods.

In some situations, it may be appropriate to conduct a sensitivity analysis to quantify the effects of nonnormality on a particular parametric statistical method (Hirsch and Slack, 1984). This is a costly approach, however, and may only be practical when very important data sets are analyzed by methods which do not have a nonparametric alternative.

³ Efficiency is defined by Bradley (1968) as "a relative term comparing the power of one test with that of a second test which acts as a standard of comparison..."

⁴ Methods to test for normality are presented in Chapter 5. They are also discussed in Harris et al. (1987) and Helsel and Hirsch (1992).

Recommendations

The best approach for dealing with potential nonnormality in ground water quality data records is to use nonparametric statistical methods. In general, parametric methods should not be used unless the data can be shown to be normally distributed.

SEASONALITY

Introduction

Seasonality is an important data characteristic which should be addressed in ground water quality DAPs. As with serial correlation, there are no clear-cut, generally accepted guidelines on how seasonality should be dealt with when analyzing water quality data. This uncertainty regarding seasonality is evident in the EPA document *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities* (U.S. EPA, 1989). Rather than providing guidelines on how to deal with it, the EPA document refers the reader to a professional statistician if seasonality is suspected.⁵

Seasonality in ground water quality variables can be caused by changes in several factors including infiltration from streams, application of fertilizers, percolation from storm events, and irrigation practices (Montgomery et al., 1987). Seasonality is the most prevalent in data collected from shallow or highly permeable aquifers (Montgomery et al., 1987).

Definition

Seasonality is the change in distribution of water quality variables which can be attributed to the time of year. A "season" can be any specified period of time but is generally one month (12 seasons per year) or three months (4 seasons per year). Seasonality may or may not occur as a consistent pattern.

Effects of Seasonality

Seasonality increases the variance of water quality data, thereby increasing the width of confidence intervals in estimation procedures

and decreasing the power of hypothesis tests. The majority of research which addresses the effects of seasonality on statistical analysis of water quality data, does so in terms of trend analysis. Seasonality can be important, however, in any statistical procedure which assumes stationary central tendency and dispersion.

Detecting Seasonality

Several approaches to checking for the presence of seasonality in water quality data are listed here:

1. Visually inspect a plot of concentrations versus time (Montgomery et al., 1987).
2. Look for annual cycles in a correlogram. Details on how to construct and interpret correlograms are given in Loftis et al. (1989).
3. Determine if there are physical factors such as a shallow or permeable aquifer which could lead to seasonality (Ward and Loftis, 1989).
4. Construct a box and whiskers plot for each season (Montgomery et al., 1987). If the boxes do not overlap, then seasonality is probably a major source of data variation (Ward and Loftis, 1989).
5. Group the data according to season and calculate the mean for each season. Then calculate the ratio of maximum to minimum mean. The higher the ratio, the greater the seasonality. This can also be done with the standard deviation. (Loftis et al., 1989).
6. Group the data according to season and conduct a Kruskal-Wallis or ANOVA test (Montgomery et al., 1987).

The above methods involve a great deal of subjectivity. For example, one person may examine a time series and say that seasonality

⁵ The EPA document (U.S. EPA, 1989) does include tentative recommendations on how to correct for seasonality when using control charts. The recommendations are summarized, however, by referring the reader to a professional statistician.

is present whereas someone else may look at the same time series and conclude that there is no significant seasonality. Even hypothesis tests (option #6) provide questionable results regarding seasonality, primarily due to the effects of sample size.⁶

Ways of Dealing With Seasonality

For both hypothesis testing and estimation, there are two commonly used approaches for dealing with seasonality in water quality data: (1) performing seasonal transformations prior to data analysis, or (2) using procedures which account for seasonality.

Some authors recommend that the presence of seasonality be verified in each individual data record prior to application of either of the two approaches for dealing with seasonality. Other authors recommend that historical data records be examined and, if seasonality appears to be a significant factor, treat all subsequent data records as being seasonal.

A selection of publications which address ways of dealing with seasonality are discussed in the following paragraphs.

Groundwater Quality: A Data Analysis Protocol (Ward et al., 1988). In their data analysis protocol, Ward et al. recommend that

special methods for dealing with seasonality be applied only if the presence of seasonality has been verified in the data record to be analyzed. Recommendations regarding ways of dealing with seasonality which are incorporated into the protocol are summarized in Table 4-1.

Notice that for trend analysis and medians comparison of independent data, the authors recommend using deseasonalized data (i.e., seasonal transformations) in conjunction with methods which do not account for seasonality. For excursion analysis (i.e., testing for a shift in concentration over a short time frame), they recommend using the original data with methods which do account for seasonality. Another important item shown in Table 4-1 is that seasonality is not a factor for medians comparison of paired data.

Techniques of Trend Analysis for Monthly Water Quality Data (Hirsch et al., 1982). This article describes the development of the Seasonal Kendall test which is the most com-

⁶ See Chapter 5 for information on the difficulties involved with using hypotheses tests.

Table 4-1. Methods for dealing with seasonality (Ward et al., 1988).

Type of Analysis	Testing Method	Seasonality	Data To Be Used
trend analysis	Kendall Tau	present	deseasonalized
		absent	original
medians comparison: independent data	Mann Whitney	present	deseasonalized
		absent	original
medians comparison: paired data	Wilcoxon signed rank	not applicable	original
excursion analysis	compute p by season	present	original
	compute overall p	absent	

monly used method for trend detection in seasonal water quality data. The Seasonal Kendall slope estimator, an estimator of trend magnitude for time series which display seasonality, is also described.⁷

The authors used Monte Carlo simulations to compare the Seasonal Kendall test to: (1) linear regression on the original data, and (2) linear regression on deseasonalized data. They found that linear regression on the original data is the best method to use if the data is known to be normal and nonseasonal. If the data is shown via statistical tests to be normal and seasonal, the best method is linear regression on deseasonalized data. They concluded, however, that the Seasonal Kendall test is the best overall method for trend detection because water quality data is usually nonnormally distributed and often seasonal. They also pointed out that tests for detecting nonnormality and seasonality are unsatisfactory particularly for small data sets. The Seasonal Kendall test has the added advantage over linear regression methods of being able to handle missing values and nondetects.

An Evaluation of Trend Detection Techniques for Use in Water Quality Monitoring Programs (Loftis et al., 1989).

Loftis et al. used Monte Carlo simulations to compare several methods of trend analysis, all of which handle seasonality in some way. Those methods are:

- Analysis of covariance (ANOCOV).
- Modified t-test.
- Kendall-tau following removal of seasonal means.
- Seasonal Kendall.
- Seasonal Kendall with serial correlation correction.⁸
- ANOCOV on ranks.
- Modified "t" on ranks.

Loftis et al. created data records with different patterns and magnitudes of seasonality in both the mean and standard deviation. They also generated records with no seasonality. Out of the seven candidate tests, the authors

recommended ANOCOV on ranks and the Seasonal Kendall test because they appeared to have the highest power.

Loftis et al. noted that ANOCOV on ranks has the benefit of being insensitive to the pattern and magnitude of seasonal change in variance. Also, ANOCOV on ranks can be conducted by using any statistical program which is capable of doing multiple linear regression. Finally, the ANOCOV method can be improved, if desired, by adding covariates to achieve better power or to model trends more accurately.

The Seasonal Kendall test has been used extensively to analyze water quality data. Also, it performs better than ANOCOV on ranks in the presence of serial correlation. For these reasons, the authors concluded that they have a slight preference for the Seasonal Kendall test.

Multivariate Tests for Trend in Water Quality (Loftis et al., 1991c). Loftis et al. recently investigated the use of multivariate methods for analyzing seasonal, serially independent water quality data.⁹ They used Monte Carlo simulations to compare the performance of the following tests applied to 10 and 20 year data records.

- A method based on work by Sen and Puri (1977) (SP).
- Multivariate analysis of variance (MANOVA).

⁷ Applications of the Seasonal Kendall test and the Seasonal Kendall slope estimator are documented in Alexander and Smith (1988), Lanfear and Alexander (1990), Walker (1991), and Lettenmaier et al. (1991).

⁸ The Seasonal Kendall test with correction for serial correlation is discussed in the section on serial correlation in this chapter.

⁹ Multivariate methods for serially correlated seasonal data are discussed in the section on serial correlation in this chapter.

- A modified version of the covariance eigenvalue test¹⁰ which assumes independence between seasons (i.e., no serial correlation) (MCE).
- A modified version of the covariance inversion test¹¹ which assumes independence between seasons (MCI).

The authors found that for normal errors, MANOVA had the highest power overall followed very closely by the SP test. The MCE and MCI tests trailed behind in terms of power.

For lognormal errors, the SP test performed as well or better than the MCE and MCI tests. MANOVA had the lowest power.

Multivariate Trend Testing of Lake Water Quality (Loftis et al., 1991b). Loftis et al. used Monte Carlo simulations to compare univariate and multivariate methods for detecting trends in seasonal water quality data. They considered only normal, serially independent errors in their study.

The authors found that MANOVA and the MCI test were "generally more powerful than their univariate counterparts applied using the Bonferroni inequality."¹² Recommendations given by Loftis et al. are as follows:

Based on these results we can make a very positive recommendation for general application of multivariate approaches... Of the two multivariate methods studied, we recommend the MCI test for routine applications because of its robust performance (as demonstrated elsewhere for rank-based methods in general) and its simplicity.

Recommendations

Some general recommendations regarding the issue of how to deal with seasonality when analyzing ground water quality data are presented here. The recommendations are based on the preceding discussion, personal judgment, and conversations with researchers and practitioners involved in ground water data analysis.

1. Determine whether seasonality is significant by examining historical data records.

Then, based on that examination, assume that all subsequent data records will either be seasonal or nonseasonal and write the data analysis protocol accordingly. The alternative to this approach is to evaluate each new data record individually for the presence of seasonality. This alternative has two major drawbacks: (1) the methods which are available to detect seasonality involve a great deal of subjectivity, particularly for small sample sizes, and (2) evaluating each new data record individually is time consuming. In addition, comparability of statistical results is enhanced if the same statistical analysis technique is used throughout the study. If data records are evaluated individually, this implies that at least two statistical methods will be employed, i.e., one which deals with seasonality and one which doesn't.

2. Avoid using seasonal transformations whenever possible. Instead, use procedures which deal with seasonality directly. Seasonal transformations change the original data and may remove important information. Also, it can be difficult to interpret estimations made from transformed data.
3. If the information goal is to compare distributions, design the sampling program so that data can be paired. Seasonality is not an issue for tests which use paired

¹⁰ The covariance eigenvalue test is discussed in the section on serial correlation in this chapter.

¹¹ The covariance inversion test is discussed in the section on serial correlation in this chapter.

¹² The "univariate counterparts" are ANOVA and the Seasonal Kendall test respectively. The Bonferroni inequality is used to control the overall significance level by performing K univariate tests at a significance level of α/K for each test.

observations such as the Wilcoxon signed rank test. In addition, tests which used paired data are more powerful than those which use independent data.

4. Use the Seasonal Kendall test if seasonality is suspected and the information goal is to detect trend. The Seasonal Kendall test has been used extensively to analyze water quality data. Also, it has been shown to have equivalent or higher power than alternative methods. The Seasonal Kendall test can easily handle missing, censored or tied data values.
5. Keep informed of new developments in the use of multivariate procedures for trend analysis of seasonal water quality data. Recent studies on lake water quality data have shown multivariate methods to be superior to multiple applications of univariate tests.

CENSORING

Introduction

Censoring occurs when chemists or data users replace numerical test results with qualitative statements. Factors which may lead to censoring include a lack of confidence in the numerical result and fear that the data may be misused.

Terminology surrounding censoring is inconsistent, confusing and controversial. There is a question of whether or not censoring should even occur because it complicates statistical analysis and causes information loss. Several methods are available, however, to statistically analyze censored data.

A data analysis protocol writer needs to address several items regarding censoring including:

- Should the laboratory be allowed to censor analytical results?
- If detection limits or codes are used for reporting data, how are they defined?

- If data is censored by the laboratory, what statistical methods should be used to analyze it?
- How should low-level data be reported by the data user?

Definition and Causes of Censoring

Censoring is the replacement of numerical laboratory measurements with qualitative explanations such as ND, <T, less than LOD, or U. In water quality analysis, censoring generally occurs at very low concentration levels where measurement reliability is in question.

Censoring can occur at two stages. The chemist may censor laboratory results in reports to the data user, and the data user may censor results in reports to management or regulatory agencies.

Lambert et al. (1991) listed the following factors which can lead to censoring by the chemist:

- The signal produced by the pollutant is too small for the instrumentation to discriminate from background noise.
- The instrumentation registers a low signal, but the chemist decides that "unpolluted" environmental samples could give a similar signal.
- A signal is registered, but certain criteria which identify the compound are not met.
- The measurement lies below a threshold set by a client or laboratory.

Censoring by the data user generally arises out of concern that individual data values will be misinterpreted. For example, a result of 0.10 ppb of TCE could cause alarm even if is well below the method detection limit. Negative results may cause people to doubt the overall integrity of the monitoring program.

If a single analyte is censored at more than one level, the data set is said to be "multiply censored." Multiple censoring can happen when: (1) various analytical methods are used for different ranges of contaminant concentration, (2) the amount of sample dilution varies, or (3) detection limits decrease over time due to improved technology (Millard and Deverel, 1988).

Definition of Detection Limits

Detection limits are boundaries set up by chemists based on criteria given to them by the data user. The criteria may include acceptable levels of Type I and Type II error. Detection limits help to describe the uncertainty associated with detecting low-level contaminants. They are an aid to interpreting the significance of single data values, whereas statistical analysis is more appropriate for interpreting a group of values.

Detection limits are the most useful if they are determined separately for each complete analytical protocol and each individual contaminant. A complete analytical protocol is a documented procedure which includes all steps in the measurement process beginning with sample preparation and ending with data presentation.

There is a great deal of disagreement among scientists on how detection limits should be determined, what they mean, and how they should be used.¹³ The confusion is evident in a

statement made by a committee which studied ACS and ASTM approaches to detection limits: "...attempts by our task force on low-level data to make a rigorous conceptual and statistical comparison of the approaches have been unsuccessful. Even similar terms are defined in different, non-comparable ways, and additional terms and concepts are used which are unique to each approach" (Brossman et al., 1988).

Most of the currently accepted definitions of detection limits have a statistical basis and can be placed into one of three categories: limits based on Type I error, limits based on Type II error, and "other" (Table 4-2). Frequently used definitions are presented in Table 4-3. Limits based on Type I and Type II errors are discussed below using the ASTM definitions.

The criterion of detection (COD), which is

¹³ The main authority on the subject of confusion over detection limits appears to be Currie (1968 and 1988).

Table 4-2. Limits for describing statistical properties of low-level data.

Organization	Limits Based On Type I Error	Limits Based On Type II Error	Other Limits	Additional References
ACS (Keith, 1991)	LOD – limit of detection, or MDL – method detection limit	RDL – reliable detection limit	LOQ – limit of quantitation	(Keith, 1983) (Porter and Ward, 1991)
ASTM (ASTM, 1990)	COD – criterion of detection	LOD – limit of detection		(Kirchmer, 1988) (Wilson, 1973)
EPA (CFR, 1991)	MDL – method detection limit		PQL – practical quantitation limit	(Glaser et al., 1981) (Keith, 1991)
Private Consultants for Love Canal Study (Lambert et al., 1991)	C _o – decision limit	D _o – detection limit		(Currie, 1968) (Currie, 1988)

Table 4-3. Definitions of limits for describing statistical properties of low-level data.

Limit	Definition	Comments
LOD*	The lowest concentration level which can be determined statistically different from a blank at a specified level of confidence (Keith, 1991).	Often set at 3σ .
RDL	The concentration level at which a detection decision is extremely likely to be made correctly (Keith, 1991).	Often set at 6σ . RDL = 2 LOD^* if $\alpha = \beta$.
LOQ	The level above which quantitative results may be obtained with a specified degree of confidence (Keith, 1991).	Often set at 10σ .
COD	The minimum quantity (analytical result) which must be observed before it can be stated that a substance has been discerned with an acceptable probability that the statement is true (ASTM, 1990).	The COD must always be accompanied by the stated probability (ASTM, 1990).
LOD**	A concentration of twice the criterion of detection when it has been decided that the risk of making a Type II error is to be equal to a Type I error (ASTM, 1990).	LOD** = 2 COD if $\alpha = \beta$.
MDL	The minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte (CFR, 1991).	Based on variability of analyte response rather than blank response (Kirchmer, 1988).
PQL	The lowest level which can be reliably achieved within specified limits of precision and accuracy during routine operation conditions (Keith, 1991).	

* as defined by ACS, ** as defined by ASTM

based on Type I error, is used to answer the question: **is the substance present?** The COD can be described in terms of a hypothesis test:

H_0 : the concentration is equal to zero
 H_a : the concentration is greater than zero
 acceptable Type I error = α
 critical value = COD

If an analytical result is greater than the COD, the substance is assumed to be present. If the result is less than the COD, there is not enough evidence to claim that the substance is present.

The location of the COD can be determined from the pdf of H_0 , the standard deviation of the analytical process, and the choice of α . If ASTM guidelines are followed, the pdf of H_0 is assumed to be normal and the standard deviation of the process is determined by one of

three analytical approaches presented in the guidelines. The relationship between the COD, α and the pdf of H_0 are shown in Figure 4-1.

The limit of detection (LOD), which is based on Type II error according to the ASTM definition, is used to answer the question: **what is the lowest concentration which can be reliably detected at a specific significance level?** The LOD can also be described in terms of a hypothesis test:

H_0 : the concentration is equal to zero
 H_a : the concentration is greater than zero
 acceptable Type I error = α
 acceptable Type II error = β , where $1 - \beta = \text{LOD}$
 critical value = COD

The location of the LOD can be determined from the pdf of H_a , the standard deviation of the analytical process, the choice of β , and the

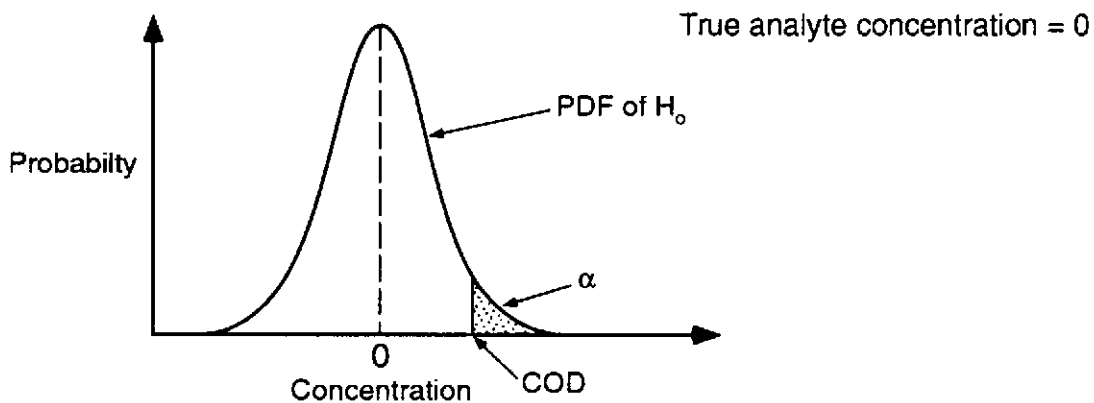


Figure 4-1. Graphic depiction of the relationship between the criterion of detection (COD), α and the PDF of H_0 .

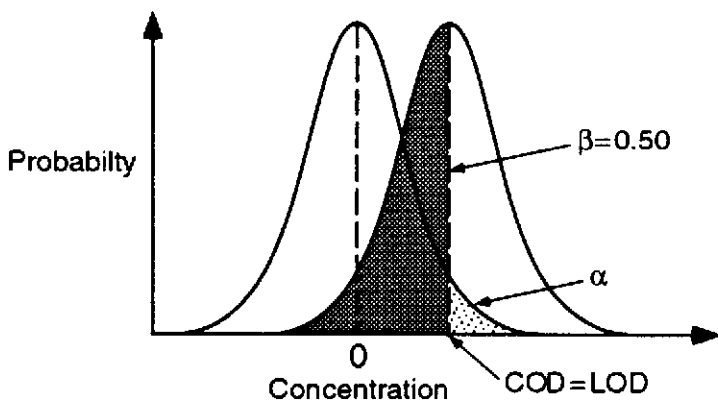


Figure 4-2. Graphic depiction of the relationship between the criterion of detection (COD), α and β when $COD = LOD$.

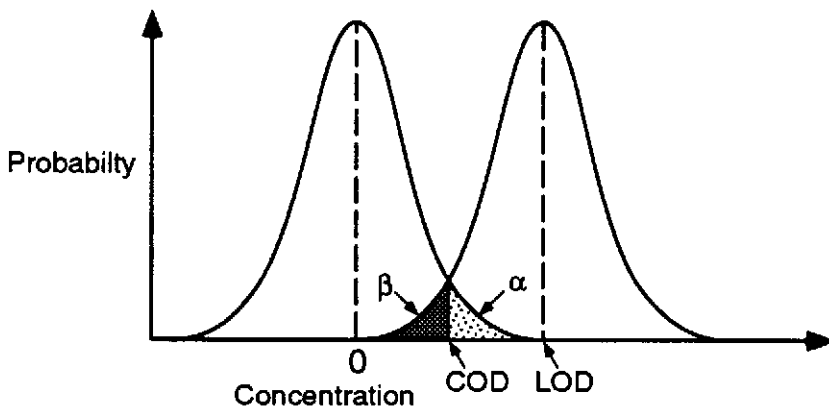


Figure 4-3. Graphic depiction of the relationship between the criterion of detection (COD), limit of detection (LOD), α and β when $\alpha = \beta$.

location of the COD. ASTM assumes that the pdf of H_a is normally distributed and that the standard deviation is independent of concentration.

The relationship between the COD, LOD, α and β is shown in Figure 4-2 for the case of $\text{LOD} = \text{COD}$, and in Figure 4-3 for the case of $\alpha = \beta$. Notice that when $\text{LOD} = \text{COD}$, β is very high (0.5), and when $\alpha = \beta$, $\text{LOD} = 2 \text{ COD}$.

The third category of detection limits (i.e., "other"), includes unique terms such as the LOQ and PQL which don't fit well into either of the first two categories. Both terms are defined in Table 4-3.

The Effects of Censoring on Statistical Analysis

Information loss is the most commonly cited effect of censoring. Porter and Ward (1991) examined the loss of information caused by measurement noise and calibration, and compared it with information loss attributable to censoring. They found that, for the conditions they studied, uncensored samples provide more information about the central tendency of a parent distribution than censored data when censoring exceeds approximately 50 percent. Gilliom et al. (1984) examined the effects of censoring on trend detection capability. They concluded that: "For all classes of data evaluated, trends were most effectively detected in uncensored data as compared to censored data even when the data censored were highly unreliable. Thus, censoring data at any concentration level may eliminate valuable information."

The effects of censoring on statistical analysis of water quality data depend on: (1) the degree of censoring, (2) the statistical method being employed, and (3) the quality of the data which has been censored.

As the degree of censoring increases, the usefulness of methods for censored data (MCDs) declines. For example, the power of tests to detect trend decreases as the percentage of censoring increases (Gilliom et al., 1984). Also, when severe (close to 50 percent or more) censoring occurs, hypothesis tests

have little power to detect differences in central tendency (Helsel, 1990). Porter et al. (1988) showed that confidence intervals for the estimates of the population mean get wider as the amount of censoring increases. Many of the MCDs simply cannot be used if censoring exceeds 50 percent (Helsel, 1990).

The effects of censoring on statistical analysis are also dependent on the statistical approach which is chosen. If simple substitution is used prior to estimation of summary statistics, the estimates may exhibit positive or negative bias depending on the substituted value. Substitution can strongly influence the results of hypothesis tests. For example, a hypothesis test may declare a significant difference if NDs are replaced by zero, but no significant difference if NDs are replaced by the MDL. Deletion of censored values can cause hypothesis tests to have "little or no meaning" (Helsel, 1990).

The third factor which influences the effects of censored data on statistical analysis is the quality of the data (Gilliom et al., 1984 and Taylor, 1988). If the analytical process used to generate the data was not in statistical control or if the bias is not predictable, then censoring may actually improve the results of statistical analysis. The obvious problem, however, is that factors which cause poor quality low-level data may also invalidate the higher concentration results.

Reporting Low-Level Data

Current recommendations which address how chemists should report their data are very much in favor of the no-censoring approach. Some of those recommendations are discussed here.

ASTM (1990) recommends that any instrument response below the COD (including negative values) be reported in conjunction with the "T" code. If no response is obtained, the "W" code should be reported along with the concentration which corresponds to the smallest increment which can be read on the analytical device.

In 1983, ACS recommended that, "Signals below 3σ should be reported as 'not detected' (ND) and the limit of detection should be given in parentheses" (Keith, 1983). In 1991, however, one of the major authors of the ACS guidelines effectively reversed this position by recommending that laboratories report all measurements (Keith, 1991). Keith (1991) also recommended that values below 3σ be flagged, and that the MDL or LOD be reported with all measurements.

The RCRA Ground Water Monitoring Technical Enforcement Guidance Document (U.S. EPA, 1985) states the following: "It is unacceptable to report only qualitative information values that were measured below a limit of detection. The technical reviewer must ensure that numerical values accompany the LT designation, so that the data are available for analysis."

The opinion that chemists should not censor data is also widely held among individuals. For example, see Lambert et al. (1991), McNichols and Davis (1988), Porter (1986),

and Porter et al. (1988). The most widely held view seems to be that chemists should report measured concentration values together with a statement of uncertainty such as detection limits or confidence intervals.

Reporting of analytical results by data users is addressed by Keith (1991). He advocates the use of censoring if "the data user determines that the data may potentially be taken and used beyond the limits defining its measurement reliability."

Methods for Statistically Analyzing Censored Data

If the decision is made to request censored data from the laboratory, the data analysis protocol writer needs to become familiar with the various methods for censored data (MCDs). Recommendations given by Helsel and Hirsch (1992) regarding MCDs are summarized in Table 4-4. Methods for estimating summary statistics of censored data sets are discussed in the following references: El-Shaarawi (1989), Gilbert (1987), Harcum

Table 4-4. Recommended methods for statistical analysis of censored data. Adapted from Helsel and Hirsch (1992).

Application	Methods
estimation of summary statistics	mean and standard deviation
	percentiles
hypothesis testing: single detection limit	compare two groups
	compare > two groups
	>50% censoring
hypothesis testing: multiple detection limits	compare two groups
	compare > two groups
regression	<20% censoring
	20% - 50% censoring
	> 50% censoring

(1990), Newman et al. (1989), and Porter and Ward (1991). Methods for analyzing multiply censored data sets are provided in Hughes and Millard (1988) and Millard and Deverel (1988). Porter (1986) is an informative source for a variety of MCDs.

Recommendations

Recommendations regarding censoring are as follows:

1. Don't allow laboratory personnel to censor data. Instead, have them report the measured concentration values together with statements of uncertainty such as detection limits or confidence intervals.
2. Use the uncensored results in statistical calculations.
3. If the potential exists for misinterpretation of low level data, consider using censoring in reports to management, regulators or the public.
4. If detection limits are used for any reason, make sure that both the determination and definition of the limits are well documented.
5. If MCDs are used, choose methods which are appropriate for the degree of censoring and the information goal, and which are robust to the assumption of normality. Also, be aware of the potential effects of censored data on the results of statistical analysis.
6. Keep informed of the latest developments in analysis and interpretation of low-level data.

SERIAL CORRELATION

Introduction

Water quality monitoring is often established in a sequential fashion through time. Samples may provide redundant information if they are taken close together relative to the

time period of interest. This redundancy of information is referred to as serial correlation. Serial correlation only becomes an issue when data is analyzed by statistical methods. It is not an intrinsic property of the data.

Serial correlation is one of the most difficult issues which a protocol writer has to deal with. There is widespread uncertainty and controversy over how serial correlation should be handled when analyzing ground water quality data. Although serial correlation has been studied extensively by theoretical statisticians, the practical application of this research is poorly developed.

Definition

Serial correlation is generally thought of as redundancy of information between adjacent observations in a time series. It can also occur between every other observation, every third observation and so on.

Positive serial correlation is characterized by the tendency for high values to follow high values and low values to follow low values. Negative serial correlation occurs when high values consistently follow low values and low values follow high ones. Only positive serial correlation will be addressed in the following discussion because water quality data is much more likely to exhibit positive than negative serial correlation.

Effects of Serial Correlation

Estimation. Confidence intervals for estimates of long-term parameters will be wider if the data is serially correlated than if it is independent. Thus, a larger sample size is needed for correlated data than independent data to achieve the same level of precision (i.e., the same confidence interval width) (Ward et al., 1990).

The phrase "long-term" refers to a characteristic estimated from data covering only a small portion of the time period of interest (Loftis et al., 1991a). For example, a person may want to estimate the annual mean concentration of selenium for the past 40 years. If the process is

assumed to be stationary, the long-term mean can be calculated by averaging samples taken at fixed intervals for a year.

The assumption of stationarity, however, is unrealistic for most water quality applications (Loftis et al., 1991a). It implies that the mean, variance, and level of serial correlation are the same throughout the time series (Gilbert, 1987). The phrase "short-term" refers to a characteristic estimated from data extending over the time period of interest (Loftis et al., 1991a). For example, a person may be interested in the mean concentration of selenium for a particular year. The short-term mean annual concentration can be calculated by averaging samples taken at fixed intervals for that year. This estimate of the mean will have a narrower confidence interval when the data is serially correlated than when it is independent (Loftis et al., 1991a).

In most references, a clear distinction is not made between long-term and short-term estimates of parameters. Sample size and variance equations which account for serial correlation are generally derived based on long-term estimates (for example, see Gilbert, 1987 p. 38). There are sample size equations which account for serial correlation when the information goal is to estimate short-term characteristics. These equations, however, are rather complex and are not widely published (Loftis et al., 1991a). The appropriateness of long-term versus short-term parameter estimates for describing water quality data is an area which needs further research.

Hypothesis Testing. Serial correlation can distort the results of any statistical test which assumes stochastic independence of observations. Most hypothesis tests, including the nonparametric versions, have this assumption.

The most commonly cited effect of positive serial correlation on hypothesis tests is inflation of the Type I error.¹⁴ In other words, for a true null hypothesis, tests will reject the hypothesis at a higher rate than the nominal significance level (usually set at 5 percent).

The end result is that the number of "false positives" are higher than they should be.

Characterizing Serial Correlation

Three facets of characterizing serial correlation are discussed here: (1) choosing a model, (2) calculating the magnitude of serial correlation, and (3) testing for significance of serial correlation.¹⁵

The simplest and most commonly used model for analyzing serial correlation in ground water quality time series is the lag 1 Markov model or AR(1). The model accounts for redundancy in the random term of adjacent observations (Harris, 1988).¹⁶

If the Markov (AR) model is chosen, serial correlation magnitude can be quantified by estimating the serial correlation coefficient, $\alpha(k)$. $\alpha(k)$ ranges from 0 to 1 for positive correlation and from 0 to -1 for negative correlation. "0" is no correlation and "k" refers to the lag.

Equations for estimating $\alpha(k)$ are given in Box and Jenkins (1970) and are applied to water quality data in Loftis et al. (1989), Montgomery et al. (1987), and Sanders et al. (1983). Harris (1988) and Close (1989) investi-

¹⁴ Inflation of the Type I error in the presence of positive serial correlation has been noted for several types of tests including: Mann-Kendall and Seasonal Kendall tests (Harcum, 1990); t, sign and Wilcoxon tests (El-Shaarawi and Damsleth, 1988); hypothesis tests based on rank correlation coefficients such as Spearman's rho and Kendall's tau (Keller-McNulty and McNulty, 1987); multiple comparison procedures (Pavur, 1988); and ANOVA (Scariano and Davenport, 1987).

¹⁵ It is commonly recommended that systematic characteristics such as seasonality and trend be removed from the time series prior to examination of serial correlation (Berryman, 1988; Montgomery and Reckhow, 1984; and Phillips et al., 1989). Careful thought should precede removal of systematic characteristics, however, so that valuable information is not discarded.

¹⁶ Additional information on model fitting can be obtained from Box and Jenkins (1970).

gated the accuracy of estimates of the lag 1 serial correlation coefficient. They both found the estimates to be too low. Harris (1988) tried using a procedure developed by Quenouille (1949) to correct for bias. She found the procedure to be unsatisfactory. Close (1989) tested a method presented in Wallis and O'Connell (1972) to correct for bias in small samples ($n < 100$). Monte Carlo simulations showed that the bias correction greatly improves the estimate of $\alpha(1)$, particularly for small sample sizes (Close, 1989).

One way to test for the significance of AR(k) serial correlation is to compare the estimated correlation coefficient to a confidence interval. If the estimate falls outside the interval, the correlation is significant (Harris et al., 1987). A graphical analog to this procedure is the correlogram. A correlogram is a plot of $\alpha(k)$ versus k for all lags k (Phillips et al., 1989). Confidence intervals are displayed as horizontal lines on the correlogram.

Harris et al. (1987) examined the power of the "confidence interval approach" to detect significant correlation of the AR(1) type. They found the power to be very low for small sample sizes. For example, the power to detect serial correlation in an AR(1) process with $\alpha(1) = 0.3$ and $n = 20$ is only 0.26. Harris et al. (1987) concluded that "roughly speaking, it is not likely that moderate amounts of serial correlation can be detected in quarterly ground-water data without at least 10 years of sampling." Consequently, serial correlation is frequently not accounted for in ground water data analysis procedures.

Another method for determining the significance of lag 1 Markov serial correlation is the rank von Neumann test (Gilbert, 1987). Harris (1988) compared the empirical power of the rank von Neumann test to that of the confidence interval approach described above. She found the rank von Neumann test to be more powerful.

Ways of Dealing With Serial Correlation

Disregard It. One option for dealing with

serial correlation is to acknowledge that it exists but don't make any adjustments in statistical calculations to account for it. The implications of this approach were introduced in the section on "effects of serial correlation."

Recall that positive serial correlation will tend to inflate the Type I error, thereby resulting in "false positives." This is generally an undesirable situation. In some instances, however, it may be appropriate to use a test which "mistakes" serial correlation for trend.

It is important to realize that the difference between serial correlation and trend is really one of scale (Loftis et al., 1991a). For example, a five year drift towards higher (or lower) values would probably be modeled as serial correlation if the period of record were 30 years, and as trend if the period of record were only five years. If short term trends (relative to the length of record) are of interest from a management point of view, it may be appropriate to "ignore" serial correlation and use an unadjusted test, accepting that some error has been introduced due to violation of the independence assumption.

Another consequence of disregarding serial correlation is that the confidence intervals on estimates of long-term parameters will be narrower than they should be. This is an undesirable situation because it may lead to overconfidence in the estimate. Conversely, the confidence intervals on estimates of short-term parameters will be wider than they should be.

Serial correlation also distorts sample size calculations. If the goal is to estimate long term characteristics, the consequences of ignoring serial correlation are underestimation of sample size. On the other hand, if the goal is to estimate short term characteristics, the presence of serial correlation will lead to overestimation of sample size. Although this is a safe scenario in that changes or differences will be easily detected, it can be expensive because unnecessary sampling and analysis is conducted.

Avoid It. A commonly used tactic for dealing with serial correlation is to avoid it by sam-

pling so infrequently that serial correlation is insignificant. Montgomery et al. (1987) analyzed 118 quarterly ground water quality data records from sites in California, Colorado and Illinois. They found that only 17 of the well records exhibited significant serial correlation at the 5 percent significance level. The general consensus seems to be that most serial correlation in ground water quality data can be avoided by taking quarterly samples (Harcum, 1990; Sara and Gibbons, 1991; and Ward et al., 1988).

The RCRA guidance document for statistical analysis of ground water data (U.S. EPA, 1989) contains a chapter on how to choose sampling intervals using the Darcy equation. Sampling intervals based on that method range from daily to monthly. Although the intent of the method is to set sampling frequencies which will result in independent data, it has been shown in the literature that monthly ground water samples can exhibit high degrees of serial correlation (Close, 1987). Given such a discrepancy, it is apparent that this subject needs further research.

If data has already been collected, an approach similar to sampling infrequently is to collapse (i.e., combine) observations until they are no longer serially correlated. For example, monthly observations could be collapsed to quarterly values. Harcum (1990) recommends using the mean to collapse data which has normally distributed errors, and the median for data which has lognormally distributed errors.

One advantage of using collapsed data values is that they have a lower variance and therefore a higher precision than single values (Ward et al, 1988). For example, quarterly averages will be more precise than a single quarterly observation (Ward et al., 1988). In general, this advantage does not outweigh the costs involved in collecting extra data.

Use Adjusted Tests. Another alternative for dealing with serial correlation is to use adjusted tests. Three types of adjusted tests are discussed here: (1) a method for using the Mann-

Whitney and Spearman's rho tests in the presence of lag 1 Markov serial correlation, (2) a modification of the Seasonal Kendall test so that it could be used with serially correlated data, and (3) extensions of the Seasonal Kendall test for use with multivariate serially correlated data.

Knowledge of the persistence structure of the time series is required in order to use any of these adjusted tests. This can present a problem because large data sets are needed to characterize serial correlation. To circumvent this problem, existing data is sometimes used to estimate serial correlation.

- **Adjustments for the Mann-Whitney and Spearman's rho tests**

Lettenmaier (1976) developed a method for using the Mann-Whitney and Spearman's rho tests in the presence of lag 1 Markov serial correlation. Montgomery and Reckhow (1984) describe Lettenmaier's method in a clear and concise manner. A summary of that description is given below:

Step 1. Calculate the test statistic using the Mann-Whitney test for linear trends or the Spearman's rho test for step trends.

Step 2. Calculate the modified critical level.

Step 3. Compare the modified critical level to the test statistic calculated in step 1. The null hypothesis is accepted or rejected under the same circumstances as for independent data.

The power can then be calculated using the concept of "equivalent independent sample size."¹⁷

¹⁷ Lettenmaier (1976) introduced the concept of "equivalent sample size" for water quality trend evaluation. [He based his research on work done by Bailey and Hammersley (1946).] For a given serially correlated time series, one can calculate the equivalent independent sample size which will provide the same amount of information as the correlated time series. For example, if weekly samples are collected (52/year), the effective independent sample size may only be 30/year due to the presence of serial correlation.

Step 1. Calculate the equivalent independent sample size.

Step 2. Calculate the trend number.

Step 3. Calculate the power.

A major drawback of the adjusted Mann-Whitney and Spearman's rho tests is that errors in estimation of the lag 1 serial correlation coefficient can have a large impact on test results. Because accurate characterization of serial correlation in water quality data is difficult (or sometimes impossible) to achieve, these adjusted tests should not be included in a data analysis protocol as routine procedures. They may be useful, however, as exploratory techniques in the initial stages of writing a DAP.

• **Modified Seasonal Kendall test**

Hirsch and Slack (1984) adjusted the Seasonal Kendall test so that it is robust against serial correlation. The modified test preserves the Type I error in the presence of serial correlation except in cases where the data records are short or have high levels of persistence (Harcum, 1990; Hirsch and Slack, 1984; and Loftis and Taylor, 1989). As mentioned previously, serial correlation causes most tests to have inflated significance levels (i.e., a high rate of false positives).

Researchers have found that the modified Seasonal Kendall test has low power in some situations. Harcum (1990) showed that if there is no serial correlation, the modified test has lower power than either the Mann-Kendall or Seasonal Kendall tests. He noted that this difference in power decreases with increasing record length. Hirsch and Slack (1984) also found that the modified test is less powerful than the original Seasonal Kendall test when there is no serial correlation. Loftis and Taylor (1989) conducted extensive simulation studies on seven different trend detection tests. They concluded that the modified test is "much less powerful than the other tests except for very large trend magnitudes and/or long data records."

Recommendations made by several authors regarding the use of the modified Seasonal

Kendall test are summarized below:

1. Hirsch and Slack (1984) recommend that the test be used for data which arise from a stationary ARMA(1,1) process, with AR parameter $F \leq 0.6$ and a record length of at least 10 years of monthly data.
2. Harcum (1990) recommends that the modified test be used on the original (monthly noncollapsed) data when there are ten or more years of data and serial correlation. He found that the significance level is preserved for ten or more years of data for "all but the highest level of serial correlation." The highest level of serial correlation he used was $\alpha(1) = 0.8$, whereas the second highest was $\alpha(1) = 0.6$.
3. Loftis and Taylor (1989) do not recommend using the modified test for routine application unless the data records are very long, "say > 20 years of quarterly data." They point out that the modified Seasonal Kendall test may "ignore" trends of moderate magnitude and duration which may be important from a management point of view. For example, a 30 year record could contain several 5 year trends, but unless a long term trend is present, the null hypothesis would not be rejected by the modified Seasonal Kendall test. Instead, the modified test would attribute the 5 year trends to serial correlation.

• **Multivariate tests**

A single multivariate test can be applied to a whole group of parameters rather than applying univariate tests to each individual parameter. Loftis et al. (1991b) recommend the use of multivariate rather than univariate tests because multivariate tests usually have superior power.

The use of multivariate testing methods to analyze water quality data is increasing. One reason for this increased popularity is that more water quality parameters are being analyzed in response to stricter regulations. Many of the parameters can be grouped together based on

their chemical compositions. Examples of such groups are common ions, nutrients and trace metals.

Lettenmaier (1988) and Loftis et al. (1991c) performed Monte Carlo simulations on three types of multivariate tests which are able to handle serial correlation. These tests are discussed below:

1. **Covariance Sum (CS) Test:** This is a multivariate extension of the univariate Seasonal Kendall test. If all trends are in the same direction, this is the most powerful of the three tests. The CS test has low power, however, when both positive and negative trends are present.
2. **Covariance Inversion (CI) Test:** This method is based on work by Dietz and Killeen (1981). Lettenmaier (1988) and Loftis et al. (1991c) showed that the CI test has very low power in most situations. Unlike the CS test, it is not negatively affected by trends of different signs.
3. **Covariance Eigenvalue (CE) Test:** The CE method was developed by Lettenmaier (1988) in an effort to improve the power of the CI test. Lettenmaier (1988) and Loftis et al. (1991c) demonstrated that the CE test generally has much better power than the CI test. The power of the CE test is not negatively affected by trends of different sign.

Recommendations

Despite the controversy and uncertainty which surrounds the issue of how to deal with serial correlation when analyzing ground water quality data, it is possible to make some general recommendations.

1. Unless there is enough evidence to justify using a more complex model, assume that serial correlation can be described by a lag 1 Markov model.
2. For sample sizes less than 100, adjust estimates of the correlation coefficient, $\alpha(1)$,

for bias using the method demonstrated in Close (1989).

3. To determine the significance of AR(k) serial correlation, use a correlogram or compare the correlation coefficient to a confidence interval. Use the rank von Neumann test to determine the significance of AR(1) serial correlation.
4. Be aware that serial correlation can occur even in quarterly data records resulting in a higher level of false positives than is indicated by the nominal significance level.
5. State the time scale of interest when estimating parameters, conducting hypothesis tests, and determining sample sizes. The importance of serial correlation is dependent on the time scale of interest.
6. Consider the possibility of sampling monthly at the beginning of a monitoring program to enable earlier characterization of serial correlation. Monthly values can be collapsed to quarterly values for the purposes of detecting differences or changes.
7. Don't use adjusted Mann-Whitney or Spearman's rho tests as routine procedures in a data analysis protocol.
8. Use the adjusted Seasonal Kendall test only for long data records with low to moderate serial correlation. Remember that the adjusted test will have lower power than most unadjusted tests when there is no serial correlation present. Also, be aware that the adjusted Seasonal Kendall test will tend to ignore short term trends which may be important from a management standpoint.
9. When serial correlation is present and it is appropriate to use multivariate tests, use the covariance sum method if trends are homogenous and the covariance eigenvalue method if trends are of different signs.

CONCLUSIONS

Data record attributes are characteristics of data which can complicate statistical analysis. Several attributes of water quality data records were reviewed in this chapter.

Multiple observations generally occur when replicate samples are collected for QA/QC purposes. If a single value is needed for statistical analysis, the multiple values should be averaged.

Outliers can either be true observations or erroneous results. If there is evidence to show that an outlier is an erroneous observation, it should be discarded. Otherwise, it should be retained and used in statistical analysis applications along with the other data. Erroneous observations can be reduced with an aggressive QA/QC program.

Changing sampling frequencies should be prevented because of the difficulties they cause for statistical analysis. Writing a data analysis protocol prior to startup of the monitoring program is an important step towards prevention.

Some statistical analysis techniques cannot

be applied to data records which have **missing values**. An effective QA/QC program can reduce the number of missing values.

Water quality data is usually **nonnormally distributed**. Tests which assume normality may have a distorted Type I error and/or low power when applied to nonnormal data. A simple solution to this problem is to use non-parametric methods.

Tests which effectively accommodate **seasonal data** have been developed. In general, it is better to use these methods than to attempt to remove seasonality.

Censoring of water quality data can cause problems with statistical analysis methods. These difficulties can be avoided by requesting the laboratory to not censor the data.

Serial correlation is probably the most difficult attribute to understand and deal with. The distinction between serial correlation and trend is poorly understood. Also, the effects of serial correlation on statistical analysis depend on whether the investigator is interested in short-term or long-term parameters, which is another poorly understood topic. Finally, there are no methods of dealing with serial correlation

Chapter 5

Statistical Analysis of Water Quality Data: Choosing Methods and Analyzing Results

INTRODUCTION

A DAP writer should know how to choose and apply statistical data analysis methods, and interpret statistical results. Choosing methods and interpreting results is the focus of this chapter. Application of statistical methods, that is, the step-by-step mechanics of conducting the procedures, is described extensively in other sources and is not covered here.

This chapter begins with an overview of the general aspects of choosing statistical methods. Separate sections are then devoted to graphical techniques, point estimation, interval estimation, and hypothesis testing. Each section includes brief summaries of some of the more commonly used procedures for analyzing water quality data. In addition, the sections on interval estimation and hypothesis testing contain detailed discussions on interpretation of statistical results.

CHOOSING STATISTICAL ANALYSIS METHODS

Introduction

Three elements which must be considered when choosing methods to statistically analyze water quality data are:

1. Monitoring information goals.
2. Data record attributes.
3. Characteristics of the proposed data analysis method.

All three elements are discussed below.

It should be emphasized that this discussion

addresses only one form of data analysis: statistics. Depending on the monitoring information goal, however, other forms of data analysis such as physical modeling or simple calculation procedures may be more appropriate. In general, statistical data analysis uses the laws of probability in conjunction with information regarding the random nature of water quality variables to provide an understanding of current water quality conditions.

Monitoring Information Goals

The first step in choosing a statistical data analysis method is to decide on a general statistical approach which matches the monitoring information goal. Several possible statistical approaches for analyzing water quality data are listed here:

- Compare two dependent groups.
- Compare two independent groups.
- Compare more than two dependent groups.
- Compare more than two independent groups.
- Determine the correlation between two continuous variables.
- Examine the relationship between two continuous variables.
- Observe and/or quantify behavior over time.
- Compare categorical data.
- Examine the relationship between continuous and categorical data.
- Estimate population distributional characteristics (e.g., mean, standard deviation).

- Verify frequency distribution assumptions (e.g., normality, equal variances).
- Estimate the probability that a single data value comes from a specific population.
- Estimate the probability that an interval contains the population value.
- Examine dependence structures in a data record (e.g., seasonality, serial correlation).

Once a statistical approach is selected, a specific method can be chosen based on data record attributes and the characteristics of candidate methods.

Data Record Attributes

The importance of considering data record attributes when choosing methods to statistically analyze water quality data is widely recognized. For example, section 25-8-204.5 of the Colorado Water Quality Control Act states: "In establishing water quality standards using statistical methodologies or in requiring the use of statistical methodologies for permit or enforcement purposes, statistical methodologies used must be based on assumptions that are compatible with the water quality data" (CDOH, 1992).

The following data record attributes were discussed in Chapter 4:

- multiple observations
- outliers
- changing sampling frequencies
- missing values
- nonnormality
- seasonality
- censoring
- serial correlation

Each attribute will be briefly addressed here in terms of its relationship to choosing statistical analysis methods. More detailed information is presented in Chapter 4.

In the case of multiple observations and changing sampling frequencies, the data should be adjusted to match the statistical method rather than vice versa. Multiple values should be averaged if a single value is needed for statistical analysis. Data which was collected with changing sampling frequencies should

be collapsed or excluded prior to using statistical methods which require equally spaced sampling intervals.

The choice of statistical method for data which contains outliers should be based on the information goal. Parametric procedures should be used if the magnitude of outliers is important. This will usually occur when dealing with units of mass. For most water quality applications, however, it is preferable to downplay the influence of outliers by using nonparametric methods. Nonparametric methods are also useful for analyzing data sets which contain missing values.

Water quality data is often nonnormally distributed. Tests which assume normality may have a distorted Type I error and/or low power when applied to nonnormal data. Water quality data analysts frequently use nonparametric methods to circumvent this problem.

Statistical tests have been developed which effectively accommodate seasonal data. In general, it is better to use these methods than to attempt to remove seasonality. The Seasonal Kendall test or ANOCOV on ranks can be used if the information goal is to detect trend. If the information goal is to compare distributions, it is best to design the sampling program so that data can be paired. Seasonality is not an issue for tests which use paired observations such as the Wilcoxon signed rank test. In addition, tests which used paired data are more powerful than those which use independent data.

Censoring can be eliminated by requesting laboratory personnel to not censor data. Not all water quality managers, however, subscribe to the no-censoring approach. If that is the case, the DAP writer must choose statistical methods which can handle the expected proportion of censored data.

When serial correlation is present and it is appropriate to use multivariate trend tests, the covariance sum method can be used for homogeneous trends and the covariance eigenvalue method can be used for trends of different signs. The most common test for analyzing serially correlated univariate data is the

adjusted Seasonal Kendall test. The adjusted test should only be used for long data records with low to moderate serial correlation. It is not appropriate to use the adjusted Mann-Whitney or Spearman's rho tests as routine procedures in a DAP because errors in estimation of the lag 1 serial correlation coefficient can have a large impact on test results.

Consideration of data record attributes when choosing a statistical analysis method can present a "chicken-and-egg" problem for the DAP writer. A major theme of DAPs is that data analysis methods are chosen prior to data collection. How then, can attributes be identified and used to choose statistical methods if the data has not yet been collected?

Nonnormality, seasonality and serial correlation can be predicted by examining historical data records from the same site or even at a similar site.

An alternative to inspecting historical data records is to write a conditional data analysis protocol. Two or more options could be provided for statistically analyzing monitoring data. After the data is collected, an option could then be chosen based on observed data characteristics. There are two problems with this approach: (1) subjectivity and complexity of the protocol is increased, and (2) ground water quality data is often difficult to characterize from small sample sizes.

Characteristics of the Data Analysis Method

Point Estimation. The performance of estimators is often gauged by their precision and bias (Berthouex and Hau, 1991). Precision refers to variability. A precise estimator has the ability to estimate a population parameter which is very close to the true population value from just one sample (Zar, 1984). This property is particularly important in ground water quality monitoring because repeated samplings are prohibitively expensive.

The second criterion which is commonly used to evaluate estimator performance is the degree of bias. Bias refers to inaccuracy caused by persistent error. If a large number of sam-

ples are collected from one population, estimates of population parameters will eventually converge to the true value if the estimator is unbiased. A biased estimator will be consistently too high or too low regardless of sample size. (Zar, 1984)

Hypothesis Tests. The performance of hypothesis tests is frequently gauged by true significance level and/or power (e.g., Conover, 1976; Harcum, 1990; and Loftis et al., 1989). The true significance level is determined with simulation studies. Tests are conducted on a large number of data sequences obtained from a population which meets the criteria of the null hypothesis. For example, if the purpose of the test is to detect trend, the true significance level would be determined by testing samples from a population which has no trend. The true significance level is the percentage of times that the null hypothesis is rejected (Loftis and Taylor, 1989). It is compared to the nominal significance level which is preassigned to the test and is commonly equal to 5 percent. The most desirable situation occurs when the true and nominal significance levels are equal (Montgomery and Loftis, 1987). If they are not equal, the true significance level will be unknown in actual (non-simulated) conditions. Incorrect conclusions could therefore be reached.

Some authors believe that power—the ability to detect departures from the null hypothesis—is the most important indicator of performance (e.g., Helsel and Hirsch, 1988). Parametric procedures can have low power if they are applied to nonnormal data (Helsel, 1987). This loss of power is critical from a water quality standpoint because it may result in nondetection of important differences or changes. Nonparametric procedures are often used because they are more powerful than their parametric counterparts for data which is nonnormally distributed.

The relative power of two test methods can be determined by simulation studies. Each test is applied to a large number of samples obtained from a population with a known departure from the null hypothesis. If test A

has a significantly higher proportion of detects¹⁸ than test B, then test A is said to be the most powerful (Harcum, 1990). It is not valid to compare the power of tests which have different true significance levels (Harcum, 1990).

Nonparametric Methods

Nonparametric methods for estimation and hypothesis testing are frequently chosen to statistically analyze water quality data. Several characteristics of nonparametric methods which make them practical for water quality data analysis are listed here:

- They are more powerful than their parametric counterparts for data sets which are nonnormally distributed.¹⁹
- Nonparametric procedures are resistant to outliers.
- They can be used to analyze data records which contain some censored values.
- They are usually easier to understand and apply than their parametric counterparts.
- Nonparametric methods can handle missing values.
- They can be applied to ordinal and sometimes nominal data.

References which discuss the advantages of nonparametric methods include: Daniel (1991), Gibbons (1985), Helsel (1987), and Hollander and Wolfe (1973).

Validity of Simulation Studies

Many recommendations presented in the literature regarding choice of statistical methods are based on simulation studies. These recommendations should be viewed with discretion for a couple of reasons. The researcher has total control over the conditions of the study, including such factors as distribution shape, type and magnitude of trends, and sample size. If conditions in nature do not match those of the simulation study (they never will exactly), then recommendations based on the study may not apply. Consequently, it is advantageous for DAP writers to review the conditions of the original simulation study.

Also, the criteria which researchers use to

evaluate the effectiveness of statistical methods may have flaws which are not yet realized. DAP writers should, therefore, keep informed of the latest advances in statistical analysis of water quality data and be prepared to revise their protocols accordingly.

Additional Considerations

Although information goals, data attributes and method characteristics are the most frequently cited considerations for choosing statistical methods, other factors may also be important. One factor is the time scale of interest. This can be particularly critical if the data is serially correlated (Loftis et al., 1991a).

Economics, politics and regulatory requirements commonly affect the choice of statistical data analysis methods.

Another consideration is the number of data sets to be analyzed. If only a few data sets are to be analyzed, it may be appropriate to use sophisticated statistical techniques such as multiple linear regression. Simpler methods may be preferable, however, if a large number of data sets are expected (Loftis et al., 1987).¹⁹

The importance of results can play a role in choice of statistical methods. For example, if a general awareness of water quality behavior over time is all that is needed, complex time series analysis would be inappropriate.

Any factors which influence the choice of statistical data analysis methods should be stated in the DAP.

GRAPHICAL METHODS

Introduction

Graphical methods are used by water quality data analysts for a variety of applications

¹⁸ A "detect" is a statistically significant result. The test is able to detect the departure from the null hypothesis.

¹⁹ Chapter 4 includes a detailed discussion of this topic.

such as:

1. Tools for choosing statistical analysis techniques.
2. A means to interpret statistical results.
3. "Stand-alone" data analysis methods.
4. A format for presenting information.

Application #1 should not be included as a routine procedure in a data analysis protocol because an underlying philosophy of DAPs is that statistical methods should be chosen prior to data collection. A DAP writer could, however, use graphical techniques to understand the general characteristics of water quality data which has already been collected at the site. This would provide the writer with valuable information regarding the type of statistical analysis methods which should be specified in the DAP.

Applications #2, #3 and #4 are appropriate for use in data analysis protocols. The importance of application #2, using graphical methods to interpret statistical results, is emphasized in many texts. For example, Ward et al. (1990) assert: "In general, it is recommended that a first step in data analysis would be to look at data in a graphical format in order to gain an understanding of water quality behavior which can be used in interpreting statistical results." Helsel and Hirsch (1992) make a similar statement: "Computing statistical measures without looking at a plot is an invitation to misunderstanding data..."

In some cases, data can be analyzed by graphical techniques alone (i.e., application #3). Graphical methods are particularly useful for analyzing limited data records. For example, a time series plot can be more informative and less misleading than a formal time series analysis if data records are short.

Application #4, using graphical methods to present information, will not be specifically addressed here. Almost any graphical method which is useful for analyzing data and interpreting results, however, can also be used to present information. Presentation graphics are examined in Helsel and Hirsch (1992, chapter 16), Schmid (1983), and Tufte (1983 and 1990).

A few of the most effective and practical graphical techniques are discussed herein. The following topics are addressed for each method: (1) potential applications, (2) types of information produced, (3) benefits and drawbacks, and (4) modifications. In most cases, the actual mechanics of constructing the graph are not given, but appropriate references are included. Graphical methods for analyzing multivariate data are not covered. References which address graphs for multivariate data include Hem (1985) and Helsel and Hirsch (1992).

Examples of completed graphs are provided for most techniques.²⁰ The same data set—hypothetical magnesium concentrations—was used for Figures 5-1 through 5-5. The right-skewed nature of the data is evident in all five figures.

Histograms

Histograms are a common graphical technique for displaying the frequency distribution of either discrete or continuous data.²¹ Histograms consist of several vertical bars, usually of equal width and variable height. The area of each bar is proportional to the number or fraction of data points falling into one of several categories or intervals (Helsel and Hirsch, 1992). A typical histogram is shown in Figure 5-1.

Histograms are useful for portraying the general shape and spread of sample distributions. They are not appropriate for detailed analyses.

Discrete data is better suited for analysis by histograms than is continuous data. Forcing continuous data into discrete groups can conceal important properties of the distribution. Furthermore, the visual impression of his-

²⁰ All graphs for the section on graphical methods were generated on Statgraphics® v.5.

²¹ Some authors use the term "histogram" for continuous data and "bar graphs" for discrete data (e.g., Weinberg and Goldberg, 1990).

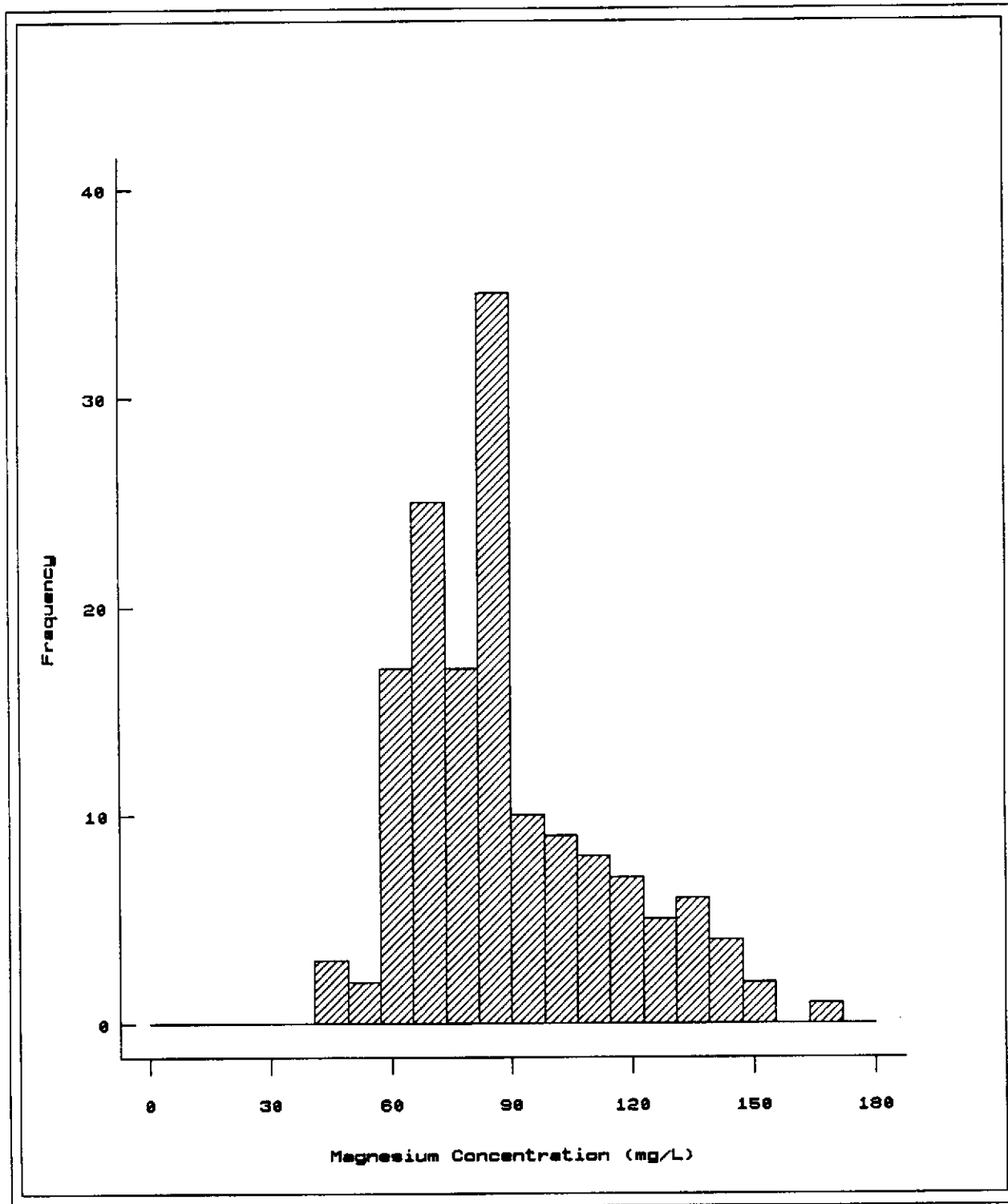


Figure 5-1. Frequency histogram.

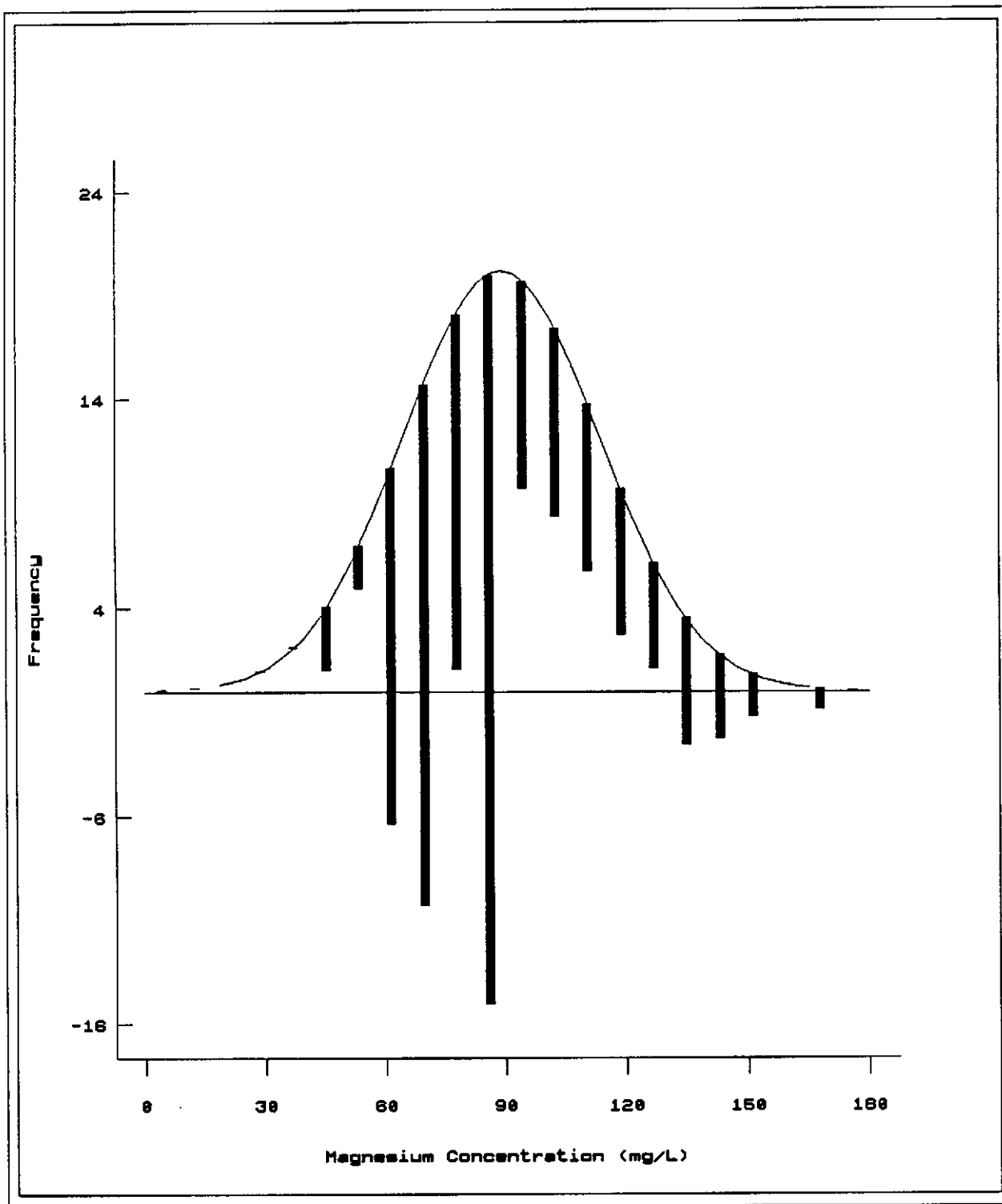


Figure 5-2. Hanging histogram.

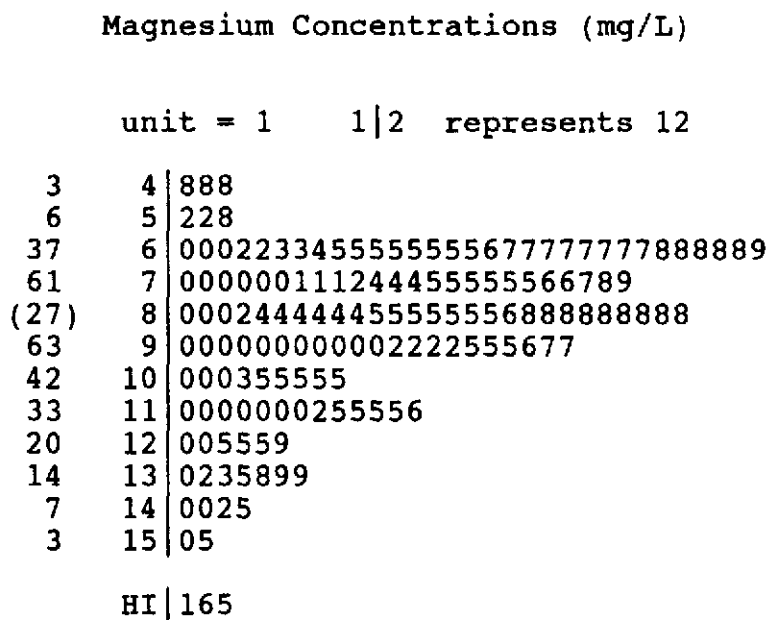


Figure 5-3. Stem and leaf diagram.

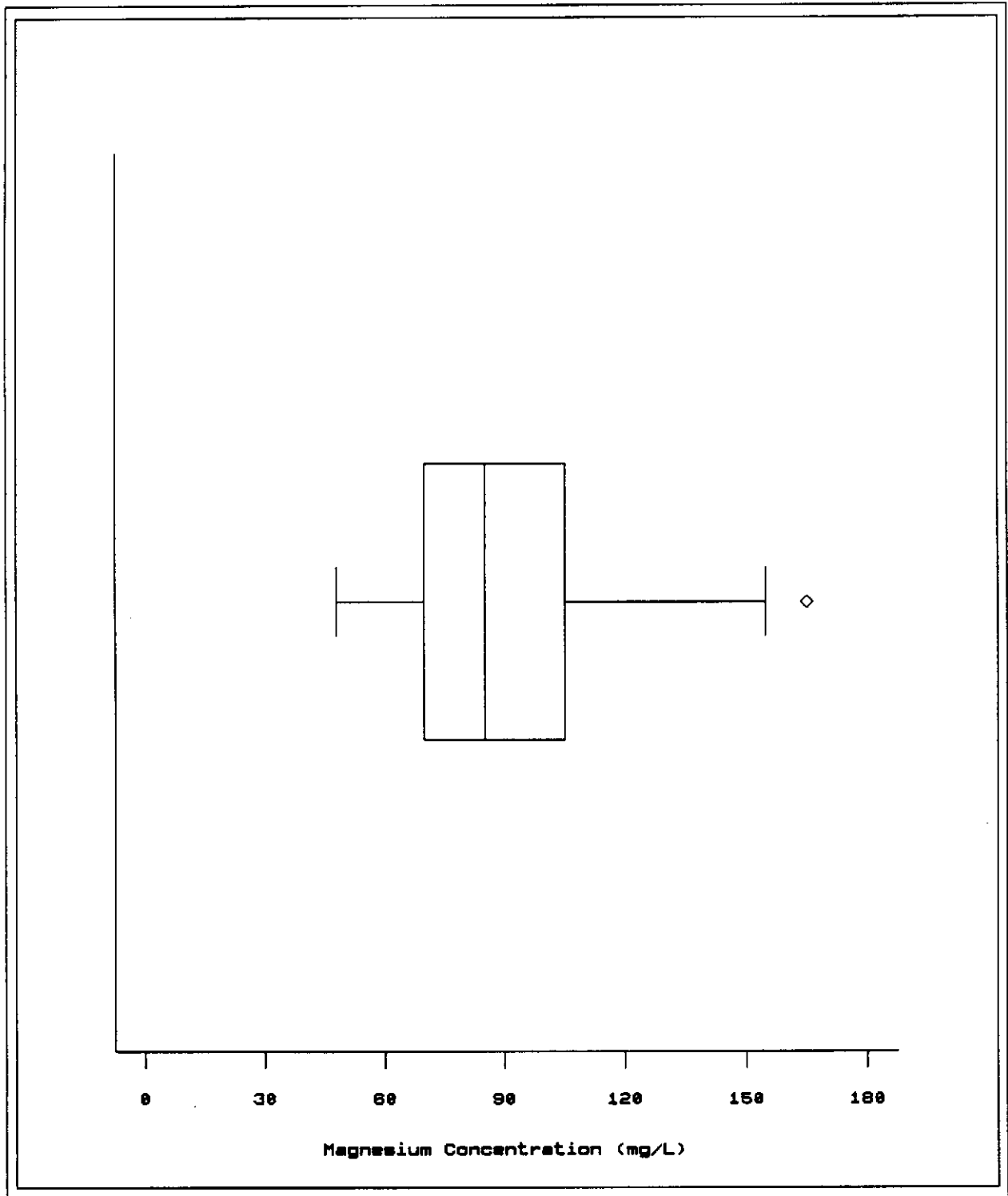


Figure 5-4. Boxplot.

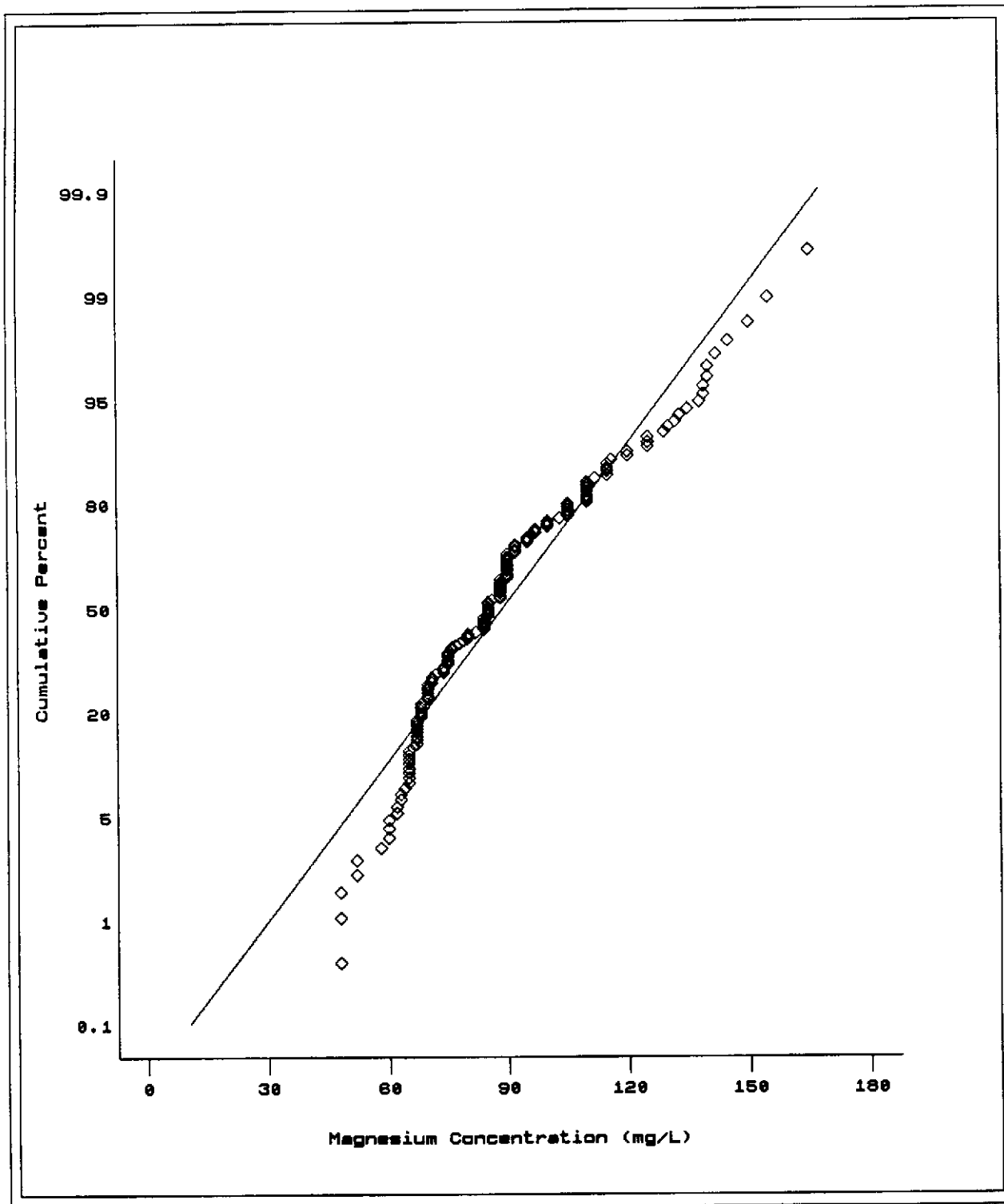


Figure 5-5. Normal probability plot.

tograms constructed from continuous data is dependent upon the number of intervals which are chosen. Discrete data often has natural groupings such as "the number of water-supply wells exceeding some critical yield grouped by geologic unit" (Helsel and Hirsch, 1992).

Histograms are the most appropriate for observing a single set of data. They are too imprecise to be of much value for comparing multiple data sets. If they are used to compare two or more groups of data, however, they should be arranged one above the other rather than side-by-side or overlapping (Helsel and Hirsch, 1992).

A variation of the histogram is the rootogram which was developed by Tukey in 1972 (Wainer and Thissen, 1981). The areas of the vertical bars in a rootogram are proportional to the square root of the counts rather than the counts themselves because the roots are usually better behaved statistically (Wainer and Thissen, 1981). Both histograms and rootograms can be hung from the best-fitting normal distribution rather than plotted from the horizontal axis (Statgraphics®, 1991 and Waiver and Thissen, 1981). If the data are normally distributed, the bottoms of the bars will be randomly scattered closely about the horizontal axis (Statgraphics®, 1991). A hanging histogram is depicted in Figure 5-2.

Stem and Leaf Diagrams

Stem and leaf diagrams resemble labeled histograms on their sides. They display more information than a histogram, however, because each data value is plotted individually. Stem and leaf diagrams are the most useful for analyzing single, small data sets. They can only be applied to continuous data. An example of a stem and leaf diagram is presented in Figure 5-3.

Wainer and Thissen (1981) describe stem and leaf diagrams as "the most important device for the analysis of small batches of numbers to appear since the t-test." Stem and leaf diagrams allow the data analyst to do the

following:

- Visually examine the distributional properties of the data.
- Calculate the range of data values.
- Check for gaps and outliers.
- Compute order statistics.

There are many different ways to construct stem and leaf diagrams. They are quite versatile and can easily be modified for a particular application. A comprehensive reference on constructing different types of stem and leaf diagrams is Tukey (1977). A simple explanation of how to produce a basic stem and leaf diagram is presented in Weinberg and Goldberg (1990).

Boxplots

Boxplots—often called box-and-whisker plots—provide a concise summary of several basic data characteristics. They are particularly useful for comparing changes in distributional characteristics of multiple data sets. Because they are highly informative, yet simple to construct and interpret, boxplots have become a popular graphical method in many disciplines. They are widely used among water quality practitioners.²²

Numerous variations of the boxplot have been developed.²³ The most commonly used version is probably the standard boxplot. A standard boxplot is shown in Figure 5-4 and its construction is briefly described here. The description is based primarily on work by Helsel and Hirsch (1992) and Velleman and Hoaglin (1981).

²² Applications of boxplots can be seen in: Barton et al. (1987), Dubrovsky and Deverel (1989), Hirsch and Gilroy (1985), Howell and El-Shaarawi (1991), Kinnamon (1993), McLeod, et al. (1983), Montgomery et al. (1987), Newell et al. (1990), and Richards (1989).

²³ References which describe variations of the boxplot include: Beckett and Gould (1987), Benjamini (1988), Tukey (1977), Velleman and Hoaglin (1981), and Ward et al. (1990). Frigge et al. (1989) discuss the variety of definitions used for boxplots in statistical software packages.

The standard boxplot is divided into several regions. The box itself contains the center 50 percent of the data (i.e., the interquartile range). Assuming the box is constructed horizontally as shown in Figure 5-4, the median is indicated as a vertical line within the box. The left end of the box is the 25th percentile and the right end is the 75th percentile. The inner fences are located at a distance of 1.5 times the width of the interquartile (IQ) range from either end of the box. The outer fences are located at a distance of 3.0 times the width of the IQ range from either end of the box. The fences are not actually marked on the boxplots, but are used to classify individual data points as described below.

Horizontal lines—sometimes called whiskers—extend from either end of the box to the outermost value within the inner fence. Data points which fall between the inner and outer fence are called “outside values” and are individually plotted (Velleman and Hoaglin, 1981). Outside values occur fewer than one time in 100 for a normal distribution (Helsel and Hirsch, 1992). Data points which fall outside the outer fence are called “far outside” values and are plotted individually with a different symbol than what is used for the outside values (Velleman and Hoaglin, 1981). Far outside values occur less than once in 300,000 times if the distribution is normal (Helsel and Hirsch, 1992).

For a single data set, boxplots provide information regarding:

- Central tendency.
- Spread.
- Skewness.
- The presence or absence of possible outliers.

For multiple data sets, boxplots indicate how data distributions change over time and between locations.

Probability Plots

Probability plots are used to observe the goodness-of-fit between a set of data and a theoretical distribution such as the normal,

lognormal or gamma distributions (Helsel and Hirsch, 1992). Although similar information can be obtained from histograms, deviations from a straight line are easier to see than deviations from a curved line (Helsel and Hirsch, 1992).

Probability plots can be produced by plotting quantiles of the sample data against quantiles of the standardized theoretical distribution. Alternatively, they can be plotted on probability paper. Statgraphics® follows the “probability paper” approach (see Figure 5-5). Details of both types of probability plot construction are presented in Helsel and Hirsch (1992).

Probability plots are most commonly used to compare data to the normal distribution. Departures from normality show up as particular patterns on a normal probability plot. If the plot is arranged so that the horizontal axis represents the variable of interest as shown in Figure 5-5, the following rule of thumb can be used: outliers at the left side of the plot will tend to fall above the linear pattern of the data, and outliers at the right side of the plot will tend to fall below the linear pattern. (The opposite is true if axes are arranged as shown in Helsel and Hirsch, 1992).

Based on this rule of thumb, the following interpretations can be made:

- Right skewed data will have a concave pattern as shown in Figure 5-5.
- Left skewed data will have a convex pattern.
- Data with heavy tails²⁴ will have an “S” shaped pattern.
- Data with light tails will have a reversed “S” shaped pattern.

These interpretations are in general agreement with du Toit et al. (1986).

Normal probability plots are the graphic analog to the probability plot correlation coefficient

²⁴ “Heavy tails” refers to the occurrence of more data points in the tails than would be expected for a normal distribution.

test (Helsel and Hirsch, 1992). The two techniques are quite effective when used in conjunction with one another. The Shapiro-Wilk test is also related to probability plots (Helsel and Hirsch, 1992).

Time Series Plots

Time series plots are simply graphs of time versus a variable such as nitrate concentration. They allow the data analyst to visually examine the time series for the following (McBride and Loftis, 1991 and Phillips et al., 1989):

- Seasonal variation.
- Trends in all or part of the data.
- Presence of extreme values.
- Homogeneity of data.
- Completeness of the data record.
- Serial correlation (to a certain extent).

Time series plots are generally used in conjunction with quantitative statistical techniques. Quantitative techniques, however, can be misleading if applied to short data records. Time series plots are an excellent "stand alone" approach for analyzing short data records, particularly if data characteristics are poorly defined.

A technique known as smoothing can be used to eliminate some of the variability in the data, thereby making trends and cycles easier to detect. Smooths are intuitively appealing because they are calculated solely from the data. A model is not assumed. They are particularly useful for making sense out of large amounts of data.²⁵ Fox et al. (1990) used a type of smooth known as LOWESS to observe trends in freshwater inflow to San Francisco Bay from the Sacramento-San Joaquin Delta.

ESTIMATION

Point Estimation

A point estimate is a single number "best guess" of some characteristic of the population. A function which is used to obtain an estimate is called an estimator. Most point estimates can be transformed into interval esti-

mates merely by adding confidence intervals.

Selected point estimators which are used for the analysis of water quality data are presented in Tables 5-1, 5-2 and 5-3. Table 5-1 includes estimators for distributional properties. Table 5-2 summarizes estimators for the difference in location based on two samples. Estimators for correlation, linear dependence and trend are presented in Table 5-3. Hypothesis tests which are commonly used in conjunction with a specific estimator are identified in the tables.

Interval Estimation

Introduction. There are three main types of interval estimation procedures:

1. Confidence intervals.
2. Prediction intervals.
3. Tolerance intervals.

Definitions and applications of each type of procedure are presented below, with references for further information.

Confidence Intervals. Confidence intervals are used routinely in the analysis of water quality data. A confidence interval is "a statement of the probability or likelihood that the interval contains the true population value" (Helsel and Hirsch, 1992). For example, a data analyst can be 95 percent confident that the true population parameter falls between the upper and lower limits of a 95 percent confidence interval.²⁶

Confidence intervals are generally used in conjunction with point estimates. Equations for calculating confidence limits have been developed for most of the estimators listed in Tables 5-1, 5-2 and 5-3. The width of the confidence interval indicates how much reliance should be placed on the point estimate. Similar information can be obtained from the yes/no results of

²⁵ See Helsel and Hirsch (1992) and Tukey (1977) for additional information on smooths.

²⁶ An excellent discussion on the interpretation of confidence intervals is presented in Jaeger (1990).

Table 5-1. Estimators of distributional properties.

Estimator or Property To Be Estimated	Application	Comments	Reference
mean	measure of location	Appropriate for computing units of mass.	Helsel and Hirsch, 1992
median	measure of location	Resistant to outliers.	" "
standard deviation	measure of spread	Unstable and inflated in the presence of outliers.	" "
interquartile range	measure of spread	Resistant to outliers.	" "
coefficient of skewness	measure of symmetry	Outliers can produce misleading results.	" "
quartile skew coefficient	measure of symmetry	Resistant to outliers. Uses only central 50 percent of data.	" "
quantiles		x_p , the p th quantile = that value below which lies 100 p % of the population. Percentile = 100 x quantile	Gilbert, 1987 Berthouex and Hau, 1991
proportions		p_{x_c} = the proportion of the population which exceeds the value x_c .	Gilbert, 1987

Table 5-2. Estimators of difference in location based on two samples.

Estimator or Property To Be Estimated	Application	Comments	Reference
difference between mean values	independent samples	Used with the two-sample t-test.	Helsel and Hirsch, 1992
Hodges-Lehmann estimator	independent samples	The median of all possible pairwise differences between the two samples. Nonparametric. Used with the rank-sum test.	Hollander and Wolfe, 1973
mean difference	dependent* samples	Applicable where differences are symmetric and normally distributed. Used with the paired t-test.	Helsel and Hirsch, 1992
Hodges-Lehmann estimator	dependent* samples	The median of all allowable pairwise averages. Nonparametric. Used with the signed-rank test.	Hollander and Wolfe, 1973
median difference	dependent* samples	Used with the sign test. Nonparametric.	Helsel and Hirsch, 1992

* "Dependent" refers to samples which can be paired.

Table 5-3. Estimators of linear dependence and monotonic trend.

Estimator or Property To Be Estimated	Application	Comments	Reference
ordinary least squares (OLS) regression*	linear dependence	The following can be estimated: slope, intercept, y given x_i , residuals for observation i , and mean square error. Assumes that residuals are normally distributed.	Helsel and Hirsch, 1992
Theil slope estimator*	linear dependence	More efficient than the OLS estimator when residuals are nonnormal. Nonparametric. Used with the Theil test for slope.	Helsel and Hirsch, 1992 Dietz, 1989 Hollander and Wolfe, 1973
Sen slope estimator	monotonic trend	Measures monotonic relationships. Nonparametric. Allows missing values. Used with the Mann-Kendall test.	Gilbert, 1987
Seasonal Kendall slope estimator	monotonic trend	Similar to the Sen slope estimator. Accounts for seasonality. Used with the Seasonal Kendall test.	Gilbert, 1987

* Can be used as a measure of trend magnitude if explanatory variable is time.

a hypothesis test accompanied by the appropriate operating characteristic curve (Natrella, 1972). Confidence intervals, however, are much easier to construct and interpret than operating characteristic curves. Another important advantage is that confidence intervals, unlike operation characteristic curves, are presented in the same units as the original observations (Natrella, 1972).

Confidence intervals are also used to detect outliers, for constructing quality control charts, and for determining sample sizes necessary to achieve a stated level of precision (Helsel and Hirsch, 1992). Control charts are discussed in the IBM data analysis protocol.

Tolerance Intervals and Prediction Intervals. Tolerance and prediction intervals are both used in ground water data analysis to determine whether a new monitoring result is consistent with background levels. These two methods allow concentration values to be considered individually while still keeping the false posi-

tive rate at a reasonable level (Gibbons, 1991a).

Contrary to what is stated in the RCRA statistical guidance document (U.S. EPA, 1989), there are important differences between tolerance intervals and prediction intervals. These differences should be considered when deciding which of the two procedures to use.

Tolerance and prediction intervals can be either one-side or two-sided. For the purposes of this discussion, one-sided intervals will be assumed because they are appropriate for determining if a ground water contaminant is higher than background levels. The following discussion is based primarily on a paper by Gibbons (1991a).

A tolerance interval states that a **given percentage of all future measurements** will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels. Therefore, a one-sided, 95 percent tolerance interval with 99 percent coverage means that a data analyst can be 95 percent confident that 99 percent of all future measure-

ments in an uncontaminated well will fall in (or below) the interval.

There are two key points in the definition of tolerance intervals. First, tolerance intervals allow a specified number of false positives. In the example given above, the analyst could be 95 percent confident that 1 of the next 100 measurements will be a false positive. Secondly, tolerance intervals are independent of the future number of measurements. The "coverage" (99 percent in the above example) refers to a specified proportion of the entire population.

A prediction interval states that **all of a given number of future measurements** will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels. Therefore, a one-sided, 95 percent prediction interval constructed for the case of $k = 20$, means that the analyst can be 95 percent confident that all of the next 20 measurements from an uncontaminated well will fall in (or below) the interval.

There are two important points in the definition of prediction intervals. The first is that no false positives are allowed. The price which is paid for this 100 percent coverage (i.e., no false positives) is an elevated false negative rate. In other words, important differences may not be detected. The second key point is that the width of the prediction interval depends on the number of future measurements for which it was constructed. The interval is narrower for fewer measurements. This feature makes prediction intervals appropriate for small monitoring programs.

Gibbons (1991a) developed a procedure using both tolerance and prediction limits which is particularly useful for situations where a large number of statistical comparisons are conducted on a regular basis. The procedure takes advantage of the strong points of both approaches and minimizes their weaknesses. Basically, Gibbons recommends that a tolerance interval be used initially, followed by the use of a prediction interval on results obtained from wells which failed the initial analysis.

Additional references on tolerance and pre-

diction intervals which may be useful to the DAP writer include: Gibbons (1987a, 1987b, 1990, 1991b), Gibbons and Baker (1991), Helsel and Hirsch (1992), Loftis et al. (1987), Mee (1990), and Sara and Gibbons (1991).

HYPOTHESIS TESTING

Introduction

Hypothesis testing is used extensively in the analysis of water quality data. Like estimation, hypothesis testing provides a means for consistency. Each analyst who uses the same data and the same test will come up with identical answers. Unlike estimation, hypothesis testing provides yes/no answers regarding the statistical significance of the evidence. "P-values" are a measure of strength of that evidence.

Hypothesis testing has historically been the basis for statistical inference. Researchers from many disciplines, however, are beginning to question the value of hypothesis tests (e.g. Graybill and Iyer, 1993). A few authors recommend abandoning them altogether in favor of interval estimation procedures. The most common opinion, though, seems to be that hypothesis tests are important tools if they are applied correctly.

The following discussion is organized around the steps involved in conducting a hypothesis test. These steps are:

1. Choose the appropriate test.
2. Establish the null and alternative hypotheses.
3. Decide on an acceptable value for α .
4. Compute the test statistic from the data.
5. Compute the p-value.
6. Reject the null hypothesis if $p \leq \alpha$.
7. Report the results.
8. Interpret the results.

Choose the Appropriate Test

The general principles involved in choosing a statistical analysis method were reviewed at the start of this chapter. Specific hypothesis tests which are used to analyze water quality data

are presented in Tables 5-4 thru 5-8. Although the tables are not exhaustive, they do contain the majority of hypothesis tests routinely used for water quality data analysis.

Establish the Null and Alternative Hypotheses

The null hypothesis, H_0 , is the hypothesis which is being tested. It is generally formulated to represent the status quo of no change or difference.²⁷ Some examples of typical null hypotheses are:

- The difference in means = 0.
- The correlation coefficient = 0.
- The slope coefficient = 0.
- The lag 1 serial correlation coefficient = 0.

The alternative hypothesis, H_a , represents a deviation from the status quo. H_a can be either one-sided or two-sided depending on the information goal. It is not valid to look at the

data and then choose a one-sided or two-sided test. Typical alternative hypotheses are:

- The difference in means $\neq 0$ (two-sided).
- The correlation coefficient $\neq 0$ (two-sided).
- The slope coefficient > 0 (one-sided).
- The lag 1 serial correlation coefficient > 0 (one-sided).

²⁷ There has been some debate over the validity of formulating the null hypothesis to represent the status quo of no change or difference. Usually, the real hypothesis of interest in water quality is whether a particular variable falls inside or outside a range of practical importance (McBride et al., 1992). A similar situation occurs in the area of bioequivalence testing. Several authors in that field have criticized traditional hypothesis testing methods and have suggested alternatives such as the equivalence test (Anderson and Hauck, 1983 and Patel and Gupta, 1984). McBride et al. (1992) discuss the use of equivalence tests to analyze water quality data.

Table 5-4. Tests for goodness of fit.

Test	Application	Comments	Reference
skewness test	normality	A simple test which works well for groundwater quality data analysis. Higher power than the chi-square test.	Harris et al., 1987 Snedecor and Cochran, 1989
Lillefors test	normality	Similar to the Kolmogorov test. Test statistic can be found graphically.	Conover, 1980
Shapiro-Wilk test	normality	A powerful test which is related to the PPCC test and normal probability plot. More complex than the skewness test.	Conover, 1980 Harris et al., 1987
probability plot correlation coefficient (PPCC) test	normality	Results can be visually illustrated on a probability plot, the graphic analog to the PPCC test.	Helsel and Hirsch, 1992
chi-square test	goodness-of-fit	Well known and versatile test. Less powerful at detecting nonnormality than most tests specifically designed for that purpose.	Harris et al., 1987 Conover, 1980
Kolmogorov test	goodness-of-fit	Preferable to chi-square test when dealing with continuous data and/or small sample sizes.	Conover, 1980 Zar, 1984

Table 5-5. Tests for location—two samples.

Test	Application	Comments	Reference
t-test	independent samples	Properties have been widely studied. Less powerful than the rank-sum test when data are nonnormally distributed.	Helsel and Hirsch, 1988 Zar, 1984
rank-sum test	independent samples	Also called the Wilcoxon rank sum, Mann-Whitney, or Wilcoxon-Mann-Whitney test. Nonparametric.	Conover, 1980 Hollander and Wolfe, 1973
paired t-test	dependent* samples	More powerful than the t-test when there is pairwise correlation. Assumes that differences are normally distributed.	Helsel and Hirsch, 1992 Zar, 1984
signed-rank test	dependent* samples	Assumes that the differences are symmetrically distributed. Nonparametric.	Conover, 1980 Hollander and Wolfe, 1973
sign test	dependent* samples	More generally applicable than the paired t-test or the signed-rank test. Easy to compute. Less powerful than the signed rank test. Nonparametric.	Conover, 1980 Gilbert, 1987

* "Dependent" refers to samples which can be paired.

Table 5-6. Tests for location—more than two samples.

Test	Application	Comments	Reference
one-factor ANOVA	independent samples	Widely used test. Assumes that all data groups have identical variances and are normally distributed.	Helsel and Hirsch, 1992 Zar, 1984
Kruskal-Wallis test	independent samples	Nonparametric equivalent of the one-factor ANOVA. Extension of the rank-sum test. Distributions of each data set are assumed to be identical in shape.	Hollander and Wolfe, 1973 Gilbert, 1987
two-factor ANOVA without replication	dependent* samples	The traditional parametric test for randomized complete block design.	Helsel and Hirsch, 1992 Zar, 1984
median aligned ranks ANOVA (MARA)	dependent* samples	An extension of the signed ranks test. Nonparametric.	Helsel and Hirsch, 1992
Friedman test	dependent* samples	The most common nonparametric test used for the randomized complete block design. An extension of the sign test.	Conover, 1980 Hollander and Wolfe, 1973
multiple comparisons procedures		Used only after an ANOVA-type test has been run which indicates a significant difference. Parametric and nonparametric versions are available.	Helsel and Hirsch, 1992

* "Dependent" refers to samples which can be paired.

Table 5-7. Tests for linear dependence and monotonic trend.

Test	Application	Comments	Reference
ordinary least squares (OLS) regression*	linear dependence	The slope and intercept can be tested for departure from zero. Assumes that residuals are normally distributed.	Helsel and Hirsch, 1992
Theil test for slope*	linear dependence	Unlike the OLS, the Theil test is not strongly affected by outliers, and residuals can be nonnormally distributed. Nonparametric.	Helsel and Hirsch, 1992 Hollander and Wolfe, 1973
Mann-Kendall test	monotonic trend	Missing values and non-detects are allowed. Nonparametric.	Gilbert, 1987 Helsel and Hirsch, 1992
Seasonal Kendall test	monotonic trend	An adaption of the Mann-Kendall test which accounts for seasonality. Widely used for water quality data analysis.	Gilbert, 1987 Hirsch et al., 1982
multivariate trend tests	monotonic trend	Multivariate tests are more powerful than repeated applications of univariate tests. MANOVA best for independent data with normal errors and Sen-Puri test best for independent data with lognormal errors. CS test best for serially dependent data with homogenous trends. CE test best for serially dependent data with trends in both directions. See Chapter 4.	Loftis et al., 1991b and 1991c

* Can be used as a test for trend if explanatory variable is time.

If a one-sided test is chosen, the hypotheses should be written in a format which is mutually exclusive. For example, if H_a is: the slope coefficient is > 0 , then H_0 should be: the slope coefficient is ≤ 0 .

Decide on an Acceptable Value for α

The significance level, α , is the probability of making a Type I error. Type I error is the rejection of the null hypothesis when it is true. If the null hypothesis is true and α is set equal to 0.05, the null hypothesis will be rejected 5 times in 100 due to chance alone.

Data analysts are free to set α at whatever value they choose. The value of α , however, inversely affects the value of β , the Type II error. Type II error is the acceptance of the null hypothesis when it is false.

The choice of α should be based on the risk

which management is willing to take of making Type I and Type II errors. A larger value of α will mean a larger number of false positives and a smaller number of false negatives.

Compute the Test Statistic from the Data

Calculation of the test statistic is dependent on the particular test which is chosen. There is usually only one way to calculate the test statistic for parametric procedures. Many nonparametric procedures, however, have three versions, all of which have different equations for determining the test statistic.

The three versions of nonparametric tests are:

1. Exact test.
2. Large sample approximation.
3. Rank transformation test.

The exact version produces exact rather than

Table 5-8. Tests for correlation, equal variance and serial dependence.

Test	Application	Comments	Reference
Pearson's r	correlation	Measures linear relationships. Not resistant to outliers. Assumes constant variance.	Helsel and Hirsch, 1992
Kendall's tau	correlation	Measures monotonic relationships. Nonparametric.	Conover, 1980 Helsel and Hirsch, 1992
Spearman's rho	correlation	Measures monotonic relationships. Nonparametric. Exact p-values should be used for $n < 20$. Spearman's ρ tends to be a larger number than Kendall's τ but the results of both tests are similar.	Conover, 1980 Helsel and Hirsch, 1992
F-test	equal variances	Severely affected by nonnormal distributions.	Zar, 1984
squared ranks test	equal variances	Nonparametric. Can be modified for use with more than two samples.	Conover, 1980
rank von Neuman test	serial dependence	Tests for the presence of lag 1 serial correlation. Nonparametric.	Gilbert, 1987 Harris, 1988
confidence interval for $r(k)$	serial dependence	Lag 1 serial correlation is significant if $r(1)$ falls outside the confidence interval.	Harris et al., 1987

approximate results. The test statistic is compared to a table of quantiles which are calculated separately for every sample size or combination of sample sizes. Exact tests should be used for small sample sizes.

The large sample approximation can be used for large sample sizes (usually > 30). This test version assumes that the test statistic follows a particular distribution such as the normal distribution. The data analyst can compare the calculated test statistic with a table for the assumed distribution, thus eliminating the need for huge tables which include exact values for large sample sizes.

Most computer programs use large sample approximations for all sample sizes. This practice can introduce significant error. If sample sizes are small, the test statistic should be obtained from the program output and compared manually to a table of exact p-values.

The third version of nonparametric tests is

the rank transformation test. Equations developed for parametric procedures are used to calculate the test statistic from ranks of the data. P-values are then determined in the same manner as for the parametric test. The rank transformation version is useful in situations where a nonparametric analog is not available for a parametric test. This can occur if a nonparametric version has never been developed, such as for the multiple-factor ANOVA, or if the nonparametric version is not included in the statistical software which is being used.

Compute the P-value

The p-value is the smallest level of significance which would have allowed the null hypothesis to be rejected (Iman and Conover, 1983). P-values are obtained from tables and are based on the value of the test statistic.

Reject the Null Hypothesis if $p \leq \alpha$

Once the p-value is determined it should be compared to the significance level, α . Reject H_0 if $p \leq \alpha$. Otherwise, accept H_0 .

Report the Results

The following information should be included when reporting the results of a hypothesis test:

- The name of the test which was used.
- The chosen significance level.
- Whether H_0 was accepted or rejected.
- The p-value.
- The sample size.
- The data themselves or the location of the data.

Some authors believe that an operating characteristic curve should also be included when reporting the results of a hypothesis test.

Interpret the Results

When interpreting the results of hypothesis tests, it is helpful to know that five parameters are interrelated in such a way that the establishment of any four of them determines the fifth (O'Brien and Shapiro, 1972). These parameters are:

1. Sample size.
2. Significance level.
3. Power.
4. Magnitude of the differences.
5. Population standard deviation.

This information is also useful for designing a sampling program (Smith and McBride, 1990). The relationship between these five parameters can be best understood by examining operating characteristic curves like the ones shown in Ward et al. (1990).

There are four possible results which can be obtained from hypothesis testing. These are shown in Table 5-9 and discussed in the following paragraphs.

If the results of a hypothesis test indicate that there is **not** a significant difference, this can mean one of two things: (1) small differences exist and they are not statistically significant, or (2) large differences exist but they are not statistically significant. #1 represents a correct decision whereas #2 is incorrect. #2 usually occurs because the sample size is too small, the power of the test is too low, or the population standard deviation is too high.

On the other hand, if results indicate that there is a significant difference, this can mean:

Table 5-9. Decisions in hypothesis testing. Adapted from Iman and Conover (1983).

		Unknown True Situation	
		H_0 is true (only small differences exist)*	H_0 is false (large differences exist)
Fail to Reject H_0	correct decision probability of correct decision = $1 - \alpha$	Type II error probability of Type II error = β	
Reject H_0	Type I error probability of Type I error = α (significance level)	correct decision probability of correct decision = $1 - \beta$ (power)	

* H_0 is never exactly true for continuous variables. Some difference always exists.

(1) large differences exist and they are statistically significant, or (2) small differences exist but they are statistically significant. Again, #1 represents a correct decision whereas #2 is incorrect. #2 can occur if sample sizes are very large or the significance level has been set at a very low level.

Another key point pertains to statistical versus practical significance. Even if the hypothesis test is correct in declaring a large difference to be statistically significant, the difference is not necessarily significant or important from a water quality standpoint. This issue has been brought up by authors in a variety of disciplines including: pharmacy (Patel and Gupta, 1984), biology (Jones and Matloff, 1986 and Perry, 1986) and public health (Brown and Mikkelsen, 1990 and Ononoff and Boden, 1987).

The p-value is a critical component of hypothesis test results. Some authors recommend that only the p-value be reported so that a reader can make his or her own decision regarding significance. For two samples of the same size, a smaller p-value indicates a higher probability that the observed effect is "real" (i.e., that it can be attributed to something other than chance). The p-value also provides an indication of the confidence with which the null hypothesis is accepted or rejected. For example, if the p-value is only slightly larger or smaller than the significance level, the decision to accept or reject H_0 is somewhat borderline.

CONCLUSIONS

Three factors to consider when choosing methods to statistically analyze water quality data are:

1. Monitoring information goals.
2. Data record attributes.
3. Characteristics of the proposed data analysis method.

Nonparametric methods for estimation and hypothesis testing are frequently chosen to statistically analyze water quality data.

Graphical methods can be included in a DAP to interpret statistical results, analyze data, or present information. Data should always be viewed in a graphical format regardless of how the data is analyzed.

A point estimate is a single number "best guess" of some characteristic of the population. Most point estimates can be transformed into interval estimates merely by adding confidence intervals.

The three main types of interval estimation procedures are confidence intervals, prediction intervals, and tolerance intervals. There are important differences between prediction and tolerance intervals which should be understood by the DAP writer.

The width of a confidence interval indicates how much reliance should be placed on the point estimate. Similar information can be obtained from the yes/no results of a hypothesis test accompanied by the appropriate operating characteristic curve. Confidence intervals, however, are easier to construct and interpret than operating characteristic curves.

Hypothesis tests are conducted using the following steps:

1. Choose the appropriate test.
2. Establish the null and alternative hypotheses.
3. Decide on an acceptable value for α .
4. Compute the test statistic from the data.
5. Compute the p-value.
6. Reject the null hypothesis if $p \leq \alpha$.
7. Report the results.
8. Interpret the results.

Although hypothesis tests have historically been the basis for statistical inference, there are many researchers who are beginning to question their value. A few authors recommend abandoning them altogether in favor of interval estimation procedures. The most common opinion, though, seems to be that hypothesis tests are important tools if they are applied correctly.

Chapter 6

A Framework for Development of Ground Water Quality Data Analysis Protocols

INTRODUCTION

The framework presented in this chapter is essentially a "how to" manual for DAP writers. Although the framework was developed for hazardous waste facilities, it can be modified for other situations such as municipal solid waste landfills. DAPs produced from the framework are to be written prior to sample collection. Existing data, however, can be used for characterization purposes or to confirm analytes.

The DAP design framework was developed with simplicity in mind. It is intended to be brief and easy to use. For this reason, detailed background material was relegated to Chapters 3, 4 and 5.

COMPONENTS OF THE FRAMEWORK

Define the Problem

The first step towards writing a data analysis protocol is to define the problem. Review all existing information on site hydrology, geology and water quality. Become familiar with current remediation activities and regulatory requirements. In some cases, a pilot study may be appropriate.

Gain Support for the Protocol

The support of users, management, and regulators is critical to the success of a DAP. Gaining support for the protocol is an ongoing process which begins prior to protocol devel-

opment. Possible approaches for obtaining support are as follows:

- Relate the advantages of a protocol. Mention that state and federal agencies, as well as private industry, are beginning to use water quality DAPs.
- Involve others in the development process by arranging meetings and circulating draft copies of the protocol.
- Include the DAP in the company's audit process.
- Back up recommendations presented in the protocol with scientific evidence from the literature, and provide references for further information.

Write the Introduction

The introduction sets the stage for the protocol. It should include purpose, scope, intended users, and limitations.

Include Procedures for Data Handling

Data handling procedures address the computerization of data, data validation, and preparation of data for analysis. If the DAP writer can verify that data handling procedures are already well documented and routinely implemented, it may not be necessary to include them in the DAP. In most cases, however, data analysis procedures are poorly defined.

State Information Goals

Identifying information goals may well be the most crucial step in writing a data analysis

protocol. It is worth spending extra time on this task. The three types of information goals are: regulatory, monitoring, and statistical. Details on identifying information goals are presented in Chapter 3.

Specify Procedures for Handling Data Record Attributes

Viable options for dealing with data record attributes are presented below. Further details are available in Chapter 4.

1. **Multiple Observations**
 - Average data.
 - Discard only if evidence is available to show that the observation is erroneous.
2. **Outliers**
 - Use nonparametric methods.
 - Discard only if evidence is available to show that the observation is erroneous.
3. **Changing Sampling Frequencies**
 - Prevent them from occurring.*
 - Collapse data.
 - Exclude data.
4. **Missing Values**
 - Prevent them from occurring.*
 - Use methods which account for them.
 - Collapse data.
5. **Nonnormality**
 - Use nonparametric methods.*
 - Confirm normality prior to using parametric methods.
 - Use transformations to obtain normality prior to using parametric methods.
6. **Seasonality**
 - Use methods which account for it.*
 - Pair data values.*
 - Disregard it and be aware of the consequences (reduction of power).
7. **Censoring**
 - Request that data from the laboratory not be censored.*
 - If data is censored, use procedures which are robust to the assumption of normality, appropriate for the infor-

mation goal, and suitable for the percentage of censored values.

8. **Serial Correlation**

- Avoid it by sampling infrequently.
- Disregard it and be aware of the consequences (usually an increase in false positives).
- Use methods which account for it.

Choose Data Analysis Methods

If statistical methods are chosen, base the choice on:

- Information goals.
- Data record attributes.
- Characteristics of the statistical method.

This topic is addressed in detail in Chapter 5. Examples of graphical techniques, tables of estimators, and tables of hypothesis tests are provided.

Describe How Statistical Results Should Be Interpreted

Always recommend that data be viewed graphically to aid in the interpretation process. Also, describe the basic information contribution of each statistical method. For example, if tolerance intervals are chosen, explain their information contribution as follows:

A tolerance interval states that a given percentage of all future measurements will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels.

If hypothesis tests are used, explain their limitations. It may help to state that five parameters—sample size, significance level, power, the magnitude of the differences, and the population standard deviation—are interrelated in such a way that the establishment of any four of them determines the fifth. This relationship can be further defined by displaying an operating characteristic curve. The interpretation of statistical analysis results is further discussed in Chapter 5.

* Preferred option for most applications.

Specify Reporting Procedures

An effective way to specify reporting procedures in a DAP is to present actual samples of how the data is to be reported. This approach was used in the case study.

In many instances, it may be appropriate to recommend the use of graphical presentation methods. Many of the graphical analysis procedures presented in Chapter 5 can also be used for presentation.

If censoring terms, such as "less than MDL", are used to report data, the terms should be clearly defined.

The statistical method should always be identified when statistical results are reported. Details on how to report the results of hypothesis tests are presented in Chapter 5.

Discuss How the Results of Statistical Analysis Should Be Utilized In Decision Making

DAPs should include explanations of how the results of statistical analyses will be used to achieve the information goals. Examples of these types of explanations are provided in the case study for information goals #2, #3 and #7.

It is important to emphasize in the information utilization section that statistical results should not be tied directly to decision making. Statistical results should be considered in conjunction with many other factors such as hydrogeology, flow patterns, and contaminant transport behavior.

Prepare a Summary Sheet for Each Information Goal

A practical strategy for organizing data analysis protocols is to prepare a summary

sheet for each information goal. This approach was very effective for the case study. Details which are specific to each information goal, such as sampling frequencies, wells to be sampled and compounds to be analyzed, can be included in the summary sheets.

State Procedures for Reviewing the Protocol

Data analysis protocols should be reviewed periodically by the DAP writer (or a worthy successor) and revised if necessary. Procedures for reviewing the protocol could address:

- Period of time between reviews.
- Special situations which would warrant a review.
- Format for documenting revisions.
- Qualifications required of the reviewer.
- Factors to consider during the review such as current research, amended regulations, and new information obtained from recent monitoring data.

Include Items to Clarify the Protocol

Numerous items can be added to data analysis protocols to enhance their clarity, such as:

- Calculation procedures (see case study).
- Glossary of terms used in protocol.
- Flow chart showing organization of protocol.
- Procedures for utilizing existing data.

SUMMARY

The main subject headings for the DAP framework are listed in Figure 6-1. The figure allows the overall structure of the framework to be seen at a glance.

- Define the problem
- Gain support for the protocol
- Write the introduction
- Include procedures for data handling
- State information goals
- Specify procedures for handling data record attributes
- Choose data analysis methods
- Describe how statistical results should be interpreted
- Specify reporting procedures
- Discuss how the results of statistical analysis should be utilized in decision making
- Prepare a summary sheet for each information goal
- State procedures for reviewing the protocol
- Include items to clarify the protocol

Figure 6-1. Main components of the framework for development of data analysis protocols for ground water quality monitoring systems.

Chapter 7

Application of the Design Framework— A Case Study

INTRODUCTION

The framework presented in Chapter 6 has been used to develop a data analysis protocol for the IBM semiconductor manufacturing facility in Hopewell Junction,²⁸ New York. The protocol will be included as an appendix to IBM's NYS (New York State) 373 permit which is currently being written. Development of the DAP represents part of IBM's continuing effort to design and operate a state-of-the-art ground water quality monitoring program.

Objectives of this chapter are as follows:

1. Provide background information which pertains to IBM's ground water monitoring program.
2. Discuss the 373 permitting process and its relationship to the DAP.
3. Describe how each step of the framework was used to write the DAP.

The IBM data analysis protocol is presented at the back of this report.

BACKGROUND

Site Description (Ward et al., 1990)

The IBM facility is one of the largest semiconductor manufacturing plants in the world. The plant began operations in 1963. Currently, there are about 4 million square feet of buildings and 10,000 employees at the site. The site covers 750 acres in a semi-rural area in southeastern New York. It is surrounded on three sides by homes and a high school.

Approximately 2 million gallons of water are used per day at the IBM site. The water comes entirely from ground water sources because there is no nearby municipal water supply. IBM treats all of its own water both before and after use.

Ground water contamination problems at the site are due primarily to the presence of chlorinated hydrocarbons including Di-, Tri- and Tetra-chloroethenes. 1,1,2,2-tetrachloroethene is the most commonly occurring compound. Chlorinated hydrocarbons are solvents which are heavily used in the manufacturing of semiconductors.

Most of the contaminants are assumed to have entered the ground water between 1963 and 1978. Sources of the contaminants have been identified as leaks from buried solvent storage tanks and transporter pipes, seepage from a former construction debris landfill and two former fire brigade training areas, and spills associated with solvent handling.

IBM has virtually eliminated further ground water contamination from the above sources. Solvents are stored in double-walled tanks which are located above ground in concrete catch basins. Many of the pipes have been replaced with double-walled pipes on trestles. Solvent handling areas are underlain with concrete, and fire training practices are environ-

²⁸ Although the IBM facility is located in Hopewell Junction, it is known as the IBM East Fishkill facility.

mentally safe. Also, 20,000 yards of material from the former construction debris landfill have been excavated and placed in a secure hazardous waste depository. The area is now filled with clean soils and capped with clay.

IBM currently operates a state-of-the-art ground water remediation program. Contaminated water is pumped from the ground and treated with air stripping and/or granulated carbon. Approximately one million gallons of water are pumped and treated daily.

Monitoring Program

Ground water at the IBM site has been monitored regularly since 1979. Major changes occurred in the monitoring system in 1982 and 1986. The initial monitoring program, the 1982 and 1986 modifications, and the monitoring system as it exists today are briefly discussed in the following paragraphs.

The initial monitoring effort which began in 1979 produced data of low quality. The data record contained many missing values and non-detects. In addition, sampling frequencies were inconsistent and data variability was high. The monitoring system design was not documented. Sampling and laboratory analysis were conducted by contractors.

IBM closely evaluated the monitoring system design and operation in 1982. The following modifications were made:

- IBM took over the field sampling and laboratory analysis.
- The monitoring system design was documented.
- Computerization of data was improved.
- Sampling points were grouped according to information goal (i.e. site perimeter, contaminated areas, and general coverage).
- Emphasis was placed on analyzing compounds associated with IBM's prior activities at the site.
- Sampling frequencies were determined based on importance of location and historical variability.
- A data verification protocol was implemented to identify outliers and confirm

their validity.

- Reporting formats were specified.
- Quality control in the laboratory and field was given high priority.
- The monitoring system design document specified that data should be periodically evaluated to substantiate the original assumptions used in the design.

Additional refinements were made to the monitoring system in 1986 to increase the information content of the water quality data.

Although the 1982 modifications resulted in higher quality data with fewer missing values, more consistent sampling frequencies and less variability, the problem of nondetects still existed. IBM decided to follow the recommendations provided by Porter (1986) and ASTM D4210-83, and not censor their data. The laboratory was asked to report all the readings together with an estimate of the measurement error.

Data variability was further reduced at some of the key wells. Dedicated pumps were installed and a bladder system was used to seal off the portion of the aquifer which was being sampled.

Data analysis revealed that additional wells were needed at some locations. New wells were added to fill in knowledge gaps, to address new concerns, and to monitor property which was added to the site.

The ground water monitoring system today includes over 550 wells and piezometers. Approximately 100 of the wells are routinely sampled for water quality. More than 2000 samples are analyzed yearly for volatile organic compounds. Priority pollutant, inorganic, radiological, and bacteriological analyses are conducted on some of the samples. Ground water level measurements are taken at all monitoring wells and piezometers.

Regulatory Requirements

The site's ground water monitoring and remediation activities have been conducted under an "Order-on-Consent" signed by IBM and the New York State Department of

Environmental Conservation (NYSDEC) in 1981. IBM is currently in the process of obtaining a permit under Part 373, Title 6 of the *New York Code of Rules and Regulations*. The Part 373 permit, once granted, will regulate ground water activities at the site.

Specifically, section 373-2.6.k.4 of 6NYCRR requires that: "In conjunction with a corrective action program, the owner or operator must establish and implement a ground-water monitoring program to demonstrate the effectiveness of the corrective action program." The regulatory basis for IBM's ground water monitoring program is further explained in the data analysis protocol at the back of this report.

373 PERMITTING PROCESS

Initial Submittal of the Ground Water Monitoring Proposal

IBM submitted the first draft of their ground water monitoring proposal on June 5, 1992. At that time, the data analysis protocol had not yet been developed.

The ground water monitoring proposal contained information on the following:

- Well locations.
- Sampling frequencies.
- Information objectives.
- Statistical approaches.
- Reporting procedures.
- Actions to be taken based on laboratory and statistical results.

Similar information was later incorporated into the DAP in a more detailed format.

The initial submittal was written primarily by IBM with minimal input from NYSDEC.

Negotiations Between NYSDEC and IBM Regarding Modifications to the Initial Ground Water Monitoring Proposal

Several meetings were held between IBM and NYSDEC to discuss modifications to the initial submittal of the ground water monitoring proposal. Those negotiations are summarized below:

- Nomenclature of wells was one of the main topics of discussion. In their initial submittal, IBM labeled several wells as "detection monitoring" wells. Although the name accurately describes the function of the wells, it conflicts with the regulatory definition of detection monitoring. To avoid future confusion over nomenclature, IBM agreed to rename the wells. On a similar note, IBM agreed to rename the "West Complex" wells as the "West Complex Detection Monitoring" wells, because they meet the regulatory definition of detection monitoring.
- Sampling frequencies were also a matter of debate. Frequencies proposed by IBM in the initial submittal ranged from monthly at the more critical wells to yearly for Appendix 33 sampling. NYSDEC agreed that Appendix 33 sampling should occur annually as per regulations. They felt, however, that the semi-annual sampling frequencies proposed by IBM should be changed to quarterly.

IBM explained their philosophy of choosing sampling frequencies which are consistent with regulatory requirements and meet information goals. They also conducted a detailed study to examine the information content of data collected at various sampling frequencies. NYSDEC agreed that it is better to have a combination of monthly, quarterly, and semi-annual frequencies which are based on factors such as risk and historical water quality, than to have across-the-board quarterly sampling.

- NYSDEC was of the opinion that a few wells should be added to the monitoring system. IBM agreed to add some of the wells.
- NYSDEC was undecided on whether information concerning wells in areas undergoing active investigation should be located in the body of the permit or in the appendices. Their most recent decision was that it should be placed in the body.

- NYSDEC wanted IBM to write a summary report of the areas under investigation. IBM agreed to write the report.

The DAP design framework was used to write a data analysis protocol during the time that NYSDEC and IBM were negotiating the ground water monitoring proposal. The protocol proved to be a useful tool for the negotiations because it is quite specific. Also, the DAP helped focus attention on information goals rather than on data quantity.

Another tool used for negotiations was a spreadsheet which compared sampling frequencies and locations initially proposed by IBM to those preferred by NYSDEC. The spreadsheet is presented in Adkins (1992).

Final Submittal of the Ground Water Monitoring Proposal

The final ground water monitoring proposal had not yet been submitted at the time this report was published. It is expected that the final proposal will be similar in format to the initial proposal. The content will reflect results of negotiations between IBM and NYSDEC.

The final ground water monitoring proposal will be placed in the body of the 373 permit, and the DAP will be included as an appendix. The DAP is quite detailed and may require periodic revision. It is much easier to revise information in the appendices than in the body of the permit.

EVALUATION OF THE FRAMEWORK

The framework provided valuable guidance for writing a data analysis protocol for the IBM facility. The combination of flexibility in the basic framework and detailed background information was quite effective. It allowed the DAP to be site specific and scientifically defensible.

Use of the framework to develop the IBM ground water quality data analysis protocol is briefly described in the following paragraphs. The discussion is structured around the main

components of the DAP framework which are listed in Figure 6-1.

Define the Problem. Extensive work had already been conducted at the IBM site to define the problem, so very little additional work was required for this first step.

Gain Support for the Protocol. Support for the protocol was gained by holding company meetings. The concept of a DAP was introduced and potential advantages were outlined. Participants included field samplers, water quality staff and managers, and laboratory personnel. Many practical suggestions were offered at the meetings.

Support was also attained by circulating rough drafts of the DAP during various stages of the development process.

Write the Introduction. Topics addressed in the introduction include:

- purpose
- advantages
- authors
- scope
- intended users
- organization

Limitations were not discussed in the introduction because they were addressed in other sections of the protocol.

Include Procedures for Data Handling. Data handling procedures are well established at the IBM site. Additional guidance and documentation in this area is unnecessary. Therefore, data handling procedures were not included in the DAP.

State Information Goals. Regulatory information goals were defined by reviewing legislation and by conversing with regulators. An outline of applicable laws and regulations was included in the protocol to guide the reader through the legislative maze.

A considerable amount of effort was devoted to formulating monitoring information

goals. The effort paid off. Regulators and water quality managers were able to get a clear picture of what information would be obtained from the monitoring program. This enabled them to identify concerns which had been overlooked or improperly emphasized. Also, monitoring information goals proved to be a convenient basis for organizing the protocol.

Details of the protocol such as sampling frequencies and data analysis methods were specified on a summary sheet for each monitoring information goal. The goals were numbered so they could be easily referenced throughout the protocol.

Detailed statistical information goals were stated on the sample calculation sheets.

Specify Procedures for Handling Data Record Attributes. Background information provided in the framework was very useful for writing this section of the protocol.

Choose Data Analysis Methods. Statistical procedures were chosen for only four of the twelve monitoring information goals. (One of the four procedures was graphical.) This rather surprising lack of emphasis on statistics can be partially attributed to the regulators' concerns. Regulators are primarily concerned that contaminated ground water is contained, collected and treated. Direct comparison with fixed limits is often all that is needed to ensure that contaminated water is being contained. Pump readings and calculations of contaminant mass provide information regarding collection and treatment.

Another reason for not using statistical methods is the potential for misinterpretation of results by those not directly involved in the monitoring program. For example, upward trends in contaminant levels may be associated

with unsuccessful remediation efforts when, in fact, they could actually be indicative of effective remediation.

In situations where statistical analysis methods were chosen, the recommendations provided by the framework were closely followed.

Describe How Statistical Results Should Be Interpreted. Interpretation of results was described for each statistical method. Difficulties related to interpretation of hypothesis tests were not an issue because estimation was chosen instead of hypothesis testing.

Specify Reporting Procedures. Reporting procedures were specified by including samples of reporting formats.

Discuss How the Results of Statistical Analysis Should Be Utilized in Decision Making. Information utilization was addressed primarily in the DAP summary sheets. It was also discussed in Appendix E which describes the data analysis methods in detail.

Prepare a Summary Sheet for Each Information Goal. Summary sheets were an effective means of organizing and conveying the myriad of details associated with each of the twelve monitoring information goals.

State Procedures for Reviewing the Protocol. Procedures were given for reviewing and updating the protocol. Recommendations provided by the DAP framework were closely followed.

Include Items to Clarify the Data Analysis Protocol. Calculation procedures were clearly presented in the protocol, and forms were provided for implementing statistical procedures.

SUMMARY

In response to rising environmental concern, the number of ground water quality monitoring systems in the United States is steadily increasing. Large amounts of money are spent collecting water samples and analyzing them in the laboratory. Despite these expenditures, however, managers often find that water quality data does not provide them with the information they need for decision making.

The inability of water quality data to provide information can usually be traced to a lack of planning in the early stages of monitoring system design. Data is frequently collected without a clear definition of program goals, data analysis procedures, reporting formats or types of decisions to be made.

It is now apparent that more information can be obtained at a lower cost if documented strategies for data analysis are developed prior to sample collection. These documented strategies, known as "data analysis protocols," help to ensure that all data obtained from monitoring programs can be translated into useful information which meets program goals.

Due to the wide variety of information needs and field conditions which are encountered in ground water monitoring, a generic data analysis protocol (DAP) would be of limited use. What is really needed is a design framework for developing DAPs which are program specific. There are currently no generally accepted guidelines on how to write water quality data analysis protocols.

The primary objective of this report is to present a framework for the development of ground water quality DAPs. Protocols developed using this framework are intended to be program specific and should be written during the initial stages of monitoring system design.

The design framework emphasizes statistical data analysis, which uses the laws of probability in conjunction with information regarding the random nature of water quality variables to provide an understanding of current water quality conditions.

Four main components of the framework include:

1. Identification of information goals.
2. Handling of data record attributes.
3. Choice of statistical analysis methods.
4. Interpretation of statistical results.

The development of DAPs should be driven by the information goals. Decisions on how to handle data attributes, which statistical methods to use, and how to interpret and report results should be consistent with information goals.

Identification of information goals is a three step process. First, regulatory information goals are identified by meeting with regulators and by reviewing regulations. Then, monitoring information goals are established. Finally, if statistical methods are used to achieve the monitoring goals, specific statistical information goals are developed.

Data record attributes are characteristics of data which can complicate statistical analysis. These include:

- multiple observations
- outliers
- changing sampling frequencies
- missing values
- nonnormality
- seasonality
- censoring
- serial correlation

Three factors to consider when choosing methods to analyze water quality data are:

1. Monitoring information goals.
2. Data record attributes.
3. Characteristics of the proposed data analysis method.

Nonparametric methods for estimation and hypothesis testing are frequently chosen to statistically analyze water quality data.

Graphical methods can be included in a DAP to interpret statistical results, analyze data, or present information. Data should always be viewed in a graphical format regardless of how the data is analyzed.

A point estimate is a single number "best guess" of some characteristic of the population. Most point estimates can be transformed into interval estimates merely by adding confidence intervals.

The three main types of interval estimation procedures are prediction intervals, tolerance intervals, and confidence intervals. There are important differences between prediction and tolerance intervals which should be understood by the DAP writer. The width of a confidence interval indicates how much reliance should be placed on the point estimate. Similar information can be obtained from the yes/no results of a hypothesis test accompanied by the appropriate operating characteristic curve. Confidence intervals, however, are easier to construct and interpret than operating characteristic curves.

Historically, hypothesis tests have been the basis for statistical inference; however, many researchers are beginning to question their value. A few authors recommend abandoning them altogether in favor of interval estimation procedures. The most common opinion,

though, seems to be that hypotheses tests are important tools if they are applied correctly.

The DAP design framework presented in Chapter 6 is essentially a "how-to" manual providing clear and concise guidelines for protocol writers.

The DAP design framework was used to develop a data analysis protocol for the IBM semiconductor manufacturing facility in Hopewell Junction, New York. The framework provided valuable guidance for writing the IBM data analysis protocol. The combination of flexibility in the basic framework and detailed background information was quite effective. It allowed the DAP to be site specific and scientifically defensible.

CONCLUSIONS

A framework for the development of ground water quality data analysis protocols is presented in this report. The framework describes the thought processes which should be followed when writing a DAP. Background information for the framework is presented in an organized and readable fashion.

Practical application of the framework is demonstrated by using it to write a DAP for the IBM semiconductor manufacturing plant in Hopewell Junction, New York. The IBM case study shows that DAPs can be an effective means for ensuring that required information is obtained from monitoring programs. The IBM data analysis protocol focuses attention on information goals, and emphasizes the information content of data rather than data quantity.

RECOMMENDATIONS FOR FURTHER WORK

1. Improve the framework which is presented here by:
 - Refining it when additional information on the statistical analysis of water quality data becomes available.

- Conducting more case studies.
 - Coding it as a computerized expert system.
2. Develop frameworks for writing protocols which address:
 - Analysis of existing data.
 - Analysis of poor quality data.
 - Information utilization.
 - Information presentation.
 3. Expand research on these topics:
 - The validity of hypothesis tests.
 - The validity of sampling quarterly to avoid serial correlation.
 - The relationship between serial correlation and time scale.
 - Estimation of short-term versus long-term parameters.
 4. Explore additional statistical data analysis methods such as equivalence tests.
 - Data visualization and spatial relationships.

CONCLUDING REMARKS

It is hoped that the framework presented in this report will elicit discussions among water quality professionals regarding standardization and use of statistical procedures. The framework is intended to be a sound starting point rather than the final word on how to write ground water quality data analysis protocols.

References

- Adkins, N.C. 1992. A framework for development of data analysis protocols for groundwater quality monitoring systems. Unpublished Ph.D. dissertation, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.
- Aldrich, J.E. 1988. Development of a hospital radiation emergency protocol. *Journal of the Canadian Association of Radiologists*. 39:137-139.
- Alexander, R.B. and R.A. Smith. 1988. Trends in lead concentrations in major U.S. rivers and their relation to historical changes in gasoline-lead consumption. *Water Resources Bulletin*. 24(3):557-569.
- Anderson, S. and W.W. Hauck. 1983. A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communication in Statistics—Theory and Methods*. 12(23):2663-2692.
- Armstrong, J.W. and D.S. Becker. 1986. Protocol standardization for Puget Sound, Washington. Estuarine and coastal management—tools of the trade. Proceedings of the Tenth National Conference of the Coastal Society. October 12-15, 1986. New Orleans, LA.
- ASTM. 1990. Standard practice for intralaboratory quality control procedures and a discussion on reporting low-level data. Designation: D 4210-89. 1990 Annual Book of Standards, Water and Environmental Technology. V. 11.01. American Society for Testing and Materials, Philadelphia, PA.
- Barton, C., E. F. Vowinkel and J. P. Nawyn. 1987. Preliminary assessment of water quality and its relation to hydrogeology and land use: Potomac-Raritan-Magothy aquifer system, New Jersey. U.S. Geological Survey Water-Resources Investigations Report 87-4023.
- Bayley, G.V. and J.M. Hammersley. 1946. The effective number of independent observations in an autocorrelated time series. *Journal of the Royal Statistical Society*. 8(1B):184-197.
- Becker, D.S. and J.W. Armstrong. 1988. Development of regionally standardized protocols for marine environmental studies. *Marine Pollution Bulletin*. 19(7):310-313.
- Beckett, S. and W. Gould. 1987. Rangefinder box plots. *The American Statistician*. 41(2):149.
- Benjamini, Y. 1988. Opening the box of a boxplot. *The American Statistician*. 42(4):257-262.
- Berryman, D., B. Bobee, D. Cluis and J. Haemmerli. 1988. Nonparametric test for trend detection in water quality time series. *Water Resources Bulletin*. 24(3):545-556.

References

- Berthouex, P.M. and W.G. Hunter. 1975. Treatment plant monitoring programs: a preliminary analysis. *Journal Water Pollution Control Federation*. 47(8):2143-2156.
- Berthouex, P.M., W.G. Hunter and L. Pallesen. 1987. Monitoring sewage treatment plants: some quality control aspects. *Journal of Quality Technology*. 10(4):139-149.
- Berthouex, P.M. and I. Hau. 1991. Difficulties related to using extreme percentiles for water quality regulations. *Research Journal of the Water Pollution Control Federation*. 63(6):873-879.
- Box, G.P. and G.M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, CA.
- Bradley, J.V. 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, NJ.
- Brinkley, K. 1987. Pronghorn hand-rearing protocol. *Animal Keepers' Forum*. 14(8):234-237.
- Brossman, M.W., G. McKenna, H. Kahn, D. King, R. Kleopfer and J.K. Taylor. 1988. Reporting low-level data for computerized databases. *Detection in Analytical Chemistry—Importance, Theory, and Practice*. L.A. Currie ed. American Chemical Society, Washington, D.C.
- Brown, D.A. 1990. EPA's resolution of the conflict between cleanup costs and the law in setting cleanup standards under Superfund. *Columbia Journal of Environmental Law*. 15(2):241-306.
- Brown, P. and E.J. Mikkelsen. 1990. *No Safe Place—Toxic Waste, Leukemia, and Community Action*. University of California Press, Berkeley, CA.
- Canter, L.W. and R.C. Knox. 1986. *Ground Water Pollution Control*. Lewis Publishers, Chelsea, MI.
- CDOH. 1992. Colorado water quality control act. Revised and reproduced by the Colorado Department of Health. November, 1990, updated August, 1992.
- CFR. 1991. Code of Federal Regulations—Protection of the Environment, 40 parts 100 to 149, revised as of July 1, 1991. Published by the Office of the Federal Register National Archives and Records Administration as a special edition of the Federal Register.
- Close, M.E. 1987. Effects of irrigation on water quality of a shallow unconfined aquifer. *Water Resources Bulletin*. 23(5):793-802.
- Close, M.E. 1989. Effect of serial correlation on ground water quality sampling frequency. *Water Resources Bulletin*. 25(3):507-515.
- Conover, W.J. and R.L. Iman. 1976. On some alternative procedures using ranks for the analysis of experimental designs. *Communication in Statistics—Theory and Methods*. A5(14):1349-1368.
- Conover, W.J. 1980. *Practical Nonparametric Statistics*. 2nd Ed. John Wiley and Sons, New York, NY.
- Cooke, S.M., B.S. Gentry, S.H. Schroeder and D.S. Berry, eds. 1987a (yearly supplements through April, 1990). *The Law of Hazardous Waste: Management, Cleanup, Liability, and Litigation*. V. 1 of 3. Matthew Bender, New York, NY.
- Cooke, S.M., B.S. Gentry, S.H. Schroeder and D.S. Berry, eds. 1987b (yearly supplements through April, 1990). *The Law of Hazardous Waste: Management, Cleanup, Liability, and Litigation*. V. 2 of 3. Matthew Bender, New York, NY.

- Cotruvo J.A., F.A. Bell, Jr., G. Ruggiero and P. Tobin. 1982. Protocol development for potable reuse criteria and standards: summary of EPA workshop. Report of Workshop Proceedings on Protocol Development; Criteria and Standards for Potable Reuse and Feasible Alternatives. SCS Eng., Reston, VA.
- Cowell, L.L. 1989. Development and validation of a ramping treadmill protocol for the on-line measurement of four aerobic parameters. Ph.D. Dissertation, University of North Texas.
- Currie, L.A. 1968. Limits for qualitative detection and quantitative determination—application to radiochemistry. *Analytical Chemistry*. 40(3):586-593.
- Currie, L.A. 1988. Detection: overview of historical, societal, and technical issues. *Detection in Analytical Chemistry—Importance, Theory, and Practice*. L.A. Currie ed. American Chemical Society, Washington, D.C. pp. 1-62.
- Daniel, W.W. 1991. *Biostatistics, a foundation for analysis in the health sciences*. 5th Ed. John Wiley and Sons, New York, NY.
- Davis, J.C. 1977. Standardization and protocols of bioassays—their role and significance for monitoring, research and regulatory usage. Proceedings 3rd Aquatic Toxicity Workshop, Halifax N.S. Nov 2-3, 1976. Environmental Protection Service Technical Report No. EPS-5-AR-77-1. Halifax, Canada. pp. 1-14.
- Dietz, E.J. 1989. Teaching regression in a nonparametrics statistics course. *The American Statistician*. 43(1):35-40.
- Dietz, E.J. and T.J. Killeen. 1981. A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*. 76:169-174.
- du Toit, S.H.C., A.G.W. Steyn and R.H. Stumpf. 1986. *Graphical Exploratory Data Analysis*. Springer-Verlag, New York, NY.
- Dubrovsky, N.M. and S.J. Deverel. 1989. Selenium in ground water of the central part of the western valley. Preliminary assessment of sources, distribution, and mobility of selenium in the San Joaquin Valley, California. U.S. Geological Survey Water-Resources Investigations Report 88-4186. pp. 35-52.
- Duke, K.M. and R.G. Merrill, Jr. 1981. Development of new bioassay protocols. *Management of Toxic Substances in Our Ecosystems—Taming the Medusa*. B.W. Cornaby, ed. Ann Arbor Science Publishers.
- Ellington, J.J., F.E. Stancil, Jr., W.D. Payne and C.D. Trusty. 1988. Interim protocol for measuring hydrolysis rate constants in aqueous solutions. U.S. Environmental Protection Agency Report No. EPA/600/3-88/014.
- El-Shaarawi, A.H. and E. Damsleth. 1988. Parametric and nonparametric tests for dependent data. *Water Resources Bulletin*. 24(3):513-519.
- El-Shaarawi, A.H. 1989. Inferences about the mean from censored water quality data. *Water Resources Research*. 25(4):685-690.
- Fox J.P., T.R. Mongan and W.J. Miller. 1990. Trends in freshwater inflow to San Francisco Bay from the Sacramento-San Joaquin Delta. *Water Resources Bulletin*. 26(1):101-116.

References

- Frigge, M., D.C. Hoaglin and B. Iglewicz. 1989. Some implementations of the boxplot. *The American Statistician*. 43(1):50-54.
- Gasith, A., K.M. Jop, K.L. Dickson, T.F. Parkerton and S.A. Kaczmarek. 1988. Protocol for the identification of toxic fractions in industrial wastewater effluents. *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971*. W.J. Adams, G.A. Chapman and W.G. Landis, eds. American Society for Testing and Materials, Philadelphia, PA. pp. 204-215.
- Gibbons, J.D. 1985. *Nonparametric Methods for Quantitative Analysis*. 2nd Ed. American Sciences Press, Columbus, OH.
- Gibbons, R.D. 1987a. Statistical prediction intervals for the evaluation of ground-water quality. *Ground Water*. 25(4):455-465.
- Gibbons, R.D. 1987b. Statistical models for the analysis of volatile organic compounds in waste disposal sites. *Ground Water*. 25(5):572-580.
- Gibbons, R.D. 1990. A general statistical procedure for ground-water detection monitoring at waste disposal facilities. *Ground Water*. 28(2):235-243.
- Gibbons, R.D. 1991a. Statistical tolerance limits for ground-water monitoring. *Ground Water*. 29(4):1991.
- Gibbons, R.D. 1991b. Some additional nonparametric prediction limits for ground-water detection monitoring at waste disposal facilities. *Ground Water*. 29(5):729-736.
- Gibbons, R.D. and J. Baker. 1991. The properties of various statistical prediction intervals for ground-water detection monitoring. *Journal of the Environmental Science and Health*. A26(4):535-553.
- Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Company, New York, NY.
- Gilliom, R.J., R.M. Hirsch and E.J. Gilroy. 1984. Effect of censoring trace-level water-quality data on trend-detection capability. *Environmental Science and Technology*. 18(7):530-535.
- Gilliom, R.J. and D.R. Helsel. 1986. Estimation of distributional parameters for censored trace level water quality data; 1. Estimation techniques *Water Resources Research*. 22(2):135-146.
- Glaser, J.A., D.L. Forest, G.D. McKee, S.A. Quave and W.L. Budde. 1981. Trace analyses for wastewaters. *Environmental Science and Technology*. 15(12):1426-1435.
- Glass, E.A. 1988. Superfund and SARA: Are There Any Defenses Left? *The Harvard Environmental Law Review*. 12(2):385-464.
- Graybill, F.A. and H.K. Iyer. 1993. *Multiple Regression Analysis—Concepts and Applications* (title is subject to change). Wadsworth Publishers, California, (not yet published).
- Greene, J.C., C.L. Bartels, W.J. Warren-Hicks, B.R. Parkhurst, G.L. Linder, S.A. Peterson and W.E. Miller. 1988. Protocols for short term toxicity screening of hazardous waste sites. U.S. Environmental Protection Agency Report No. EPA/600/3-88/029.
- Hall, R.M., Jr., T. Watson, J.J. Davidson, D.R. Case and N.S. Bryson (Crowell and Moring). 1987. *RCRA Hazardous Waste Handbook*. 7th Ed. Government Institutes, Inc., Rockville, MD.

- Hall, R.M., Jr., T. Watson, R.E. Schwartz, N.S. Bryson and R.C. Davis, Jr. (Crowell and Moring). 1990. Superfund Manual. 3rd Ed. Government Institute, Inc., Rockville, MD.
- Harcum, J.B. 1990. Water-quality data analysis protocol development. Unpublished Ph.D. Dissertation, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.
- Harris, J., J.C. Loftis and R.H. Montgomery. 1987. Statistical methods for characterizing ground-water quality. *Ground Water*. 25(2):185-193.
- Harris, J. 1988. Statistical analysis of ground water quality-interaction of deterministic and stochastic components. Unpublished Ph.D. Dissertation, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.
- Hedberg, A.M., N. Garcia, I.J. Trejus, S. Weinmann-Winkler, M.L. Gabriel and A.L. Lutz. 1988. Nutritional risk screening: development of a standardized protocol using dietetic technicians. *Journal of the American Dietetic Association*. 88(12):1553-1556.
- Helsel, D.R. 1987. Advantages of nonparametric procedures for analysis of water quality data. *Hydrological Sciences—Journal*. 32(2):179-190.
- Helsel, D.R. and R.M. Hirsch. 1988. Discussion: "Applicability of the t-test for detecting trends in water quality variables" by Robert H. Montgomery and Jim C. Loftis. *Water Resources Bulletin*. 24(1):201-204.
- Helsel, D.R. 1990. Less than obvious—statistical treatment of data below the detection limit. *Environmental Science and Technology*. 24(12):1766-1774.
- Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. Elsevier Publishers.
- Hem, J. 1985. Study and interpretation of the chemical characteristics of natural water. 3rd Ed. U.S. Geological Survey Water-Supply Paper 2254.
- Hirsch, R.M., J.R. Slack and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*. 18(1):107-121.
- Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*. 20(6):727-732.
- Hirsch, R.M. and E.J. Gilroy. 1985. Detectability of step trends in the rate of atmospheric deposition of sulfate. *Water Resources Bulletin*. 21(5):773-784.
- Hirsch, R.M. 1988. Statistical methods and sampling design for estimating step trends in surface-water quality. *Water Resources Bulletin*. 24(3):493-503.
- Hirsch, R.M., R.B. Alexander and R.A. Smith. 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research*. 27(5):803-813.
- Hockman, K.K. and J.M. Lucas. 1987. Variability reduction through subvessel CUSUM control. *Journal of Quality Technology*. 19(3):113-121.
- Hodges, J.L. and E.L. Lehmann. 1963. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*. 34:598-611.

References

- Hollander, M. and D.A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York, NY.
- Howell, G., and A. H. El-Shaarawi. 1991. An overview of the acidification of lakes in Atlantic Canada. *Environmental Monitoring and Assessment*. 17:323-338.
- Hughes, J.P. and S.P. Millard. 1988. A tau-like test for trend in the presence of multiple censoring points. *Water Resources Bulletin*. 24(3):521-531.
- Hunter, W.G. 1981. The practice of statistics: the real world is an idea whose time has come. *The American Statistician*. 35(2).
- Iman, R.L. and W.J. Conover. 1983. *A Modern Approach to Statistics*. John Wiley and Sons, New York, NY.
- Jones, D. and N. Matloff. 1986. Statistical hypothesis testing in biology; a contradiction in terms. *FORUM: Journal of Economic Entomology*. 79(5):1156-1160.
- Jorgensen, E.P., ed., (Sierra Club Legal Defense Fund). 1989. *The Poisoned Well*. Island Press, Washington D.C.
- Keith, L.H. 1991. *Environmental Sampling and Analysis: A Practical Guide*. Lewis Publishers, Chelsea, MI.
- Keith, L.H., W. Crummett, J. Deegan, Jr., R.A. Libby, J.K. Taylor and G. Wentler. 1983. Principles of environmental analysis. *Analytical Chemistry*. 55(14):2210-2218.
- Keller-McNulty, S. and M. McNulty. 1987. The independent pairs assumption in hypothesis tests based on rank correlation coefficients. *The American Statistician*. 41(1):40-41.
- Kinnamon, D.E. 1993. Evaluation of ground water migration using areal differences in water quality. Unpublished M.S. Thesis, Department of Environmental Health, Colorado State University, Fort Collins, CO.
- Kirchmer, C.J. 1988. Estimation of detection limits for environmental analytical procedures—a tutorial. *Detection in Analytical Chemistry—Importance, Theory, and Practice*. L.A. Currie ed. American Chemical Society, Washington, D.C.
- Lambert, D., B. Peterson and I. Terpenning. 1991. Nondetects, detection limits, and the probability of detection. *Journal of the American Statistical Association*. 86(414):266-277.
- Lanfear, K.J. and R.B. Alexander. 1990. Methodology to derive water-quality trends for use by the National Water Summary Program of the U.S. Geological Survey. U.S. Geological Survey Open-File Report 90-359, Reston, VA.
- Lettenmaier, D.P. 1976. Detection of trends in water quality data from records with dependent observations. *Water Resources Research*. 12(5):1037-1046.
- Lettenmaier, D.P. 1988. Multivariate nonparametric tests for trend in water quality. *Water Resources Bulletin*. 24(3):505-512.
- Lettenmaier, D.P., E.R. Hooper, C. Wagoner and K.B. Faris. 1991. Trends in stream quality in the continental United States. *Water Resources Research*. 27(3):327-339.

- Lipson, S.F. and P.T. Ellison. 1987. Development of protocols for collection of saliva samples under field conditions. (abstract only). *American Journal of Physical Anthropology*. 72(2):226.
- Loftis, J.C., P.S. Porter and G. Settembre. 1987. Statistical analysis of industrial wastewater monitoring data. *Journal of the Water Pollution Control Federation*. 59(3):145-151.
- Loftis, J.C. and C.H. Taylor. 1989. Testing for trend in water quality data. *Proceedings International Symposium on the Design of Water Quality Information Systems*, held in Fort Collins, CO on June 7-9, 1989. R.C. Ward, J.C. Loftis and G.B. McBride, eds. pp. 461-467. Information Series No. 61, Colorado Water Resources Research Institute, Fort Collins, CO.
- Loftis, J.C., R.C. Ward, R.D. Phillips and C.H. Taylor. 1989. An evaluation of trend detection techniques for use in water quality monitoring programs. U.S. Environmental Protection Agency Report No. EPA/600/3-89/037.
- Loftis, J.C., G.B. McBride and J.C. Ellis. 1991a. Considerations of scale in water quality monitoring and data analysis. *Water Resources Bulletin*. 27(2):1-10.
- Loftis, J.C., C.H. Taylor, A.D. Newell and P.L. Chapman. 1991b. Multivariate trend testing of lake water quality. *Water Resources Bulletin*. 27(3):461-473.
- Loftis, J.C., C.H. Taylor and P.L. Chapman. 1991c. Multivariate tests for trend in water quality. *Water Resources Research*. 27(7):1419-1429.
- Lucas, J.M., 1982. Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology*. 14(2):51-59.
- Marmorek D.R., D.P. Bernard, C.H.R. Wedeles, G. Sutherland, J.A. Malanchuk and W.E. Fallon. 1989. A protocol for determining lake acidification pathways. *Water, Air, and Soil Pollution*. 44:235-257.
- Marriott, F.H.C. 1990. *A Dictionary of Statistical Terms*. Longman Scientific and Technical, Essex, England, and John Wiley and Sons, Inc., New York, NY.
- Maxwell R.J. and W.N. Marmer. 1983. Systematic protocol for the accumulation of fatty acid data from multiple tissue samples: tissue handling, lipid extraction and class separation, and capillary and chromatographic analysis. *Lipids*. 18(7).
- McBride, G. and J. Loftis. 1991. *Statistical Methods for Water Quality Sampling Programmes—notes for the short course*. February 25-28 and March 11-23, 1991. Organized by Water Quality Centre, DSIR, Hamilton, New Zealand.
- McBride, G.B., J.C. Loftis and N.C. Adkins. 1992. What do significance tests really tell us about the environment? Submitted to *Environmental Management*.
- Mcleod, A.I., K.W. Hipel and F. Comancho. 1983. Trend assessment of water quality time series. *Water Resources Bulletin*. 19(4):537-547.
- McNichols, R.J. and C.B. Davis. 1988. Statistical issues and problems in ground water detection monitoring at hazardous waste facilities. *Ground Water Monitoring Review*. Fall, 1988.
- Mee, R.W. 1990. Confidence intervals for probabilities and tolerance regions based on a generalization of the Mann-Whitney statistic. *Journal of the American Statistical Association*. 85(411):793-800.

References

- Mielke, P.W. and K.J. Berry. 1976. Multi-response permutation procedures for a priori classification. *Communications in Statistics—Theory and Methods*. A5(14):1409-1424.
- Millard, S.P. and S.J. Deverel. 1988. Nonparametric statistical methods for comparing two sites based on data with multiple detection limits. *Water Resources Research*. 24(12):2087-2098.
- Misuriello, H.P. 1987. A uniform procedure for the development and dissemination of building monitoring protocols. *American Society of Heating, Refrigeration, and Air Conditioning Engineers—Transactions*. 93(1):1619-1629.
- Montgomery, R.H. and K.H. Reckhow. 1984. Techniques for detecting trends in lake water quality. *Water Resources Bulletin*. 20(1):43-52.
- Montgomery, R.H. and J.C. Loftis. 1987. Applicability of the t-test for detecting trends in water quality variables. *Water Resources Bulletin*. 23(4):653-662.
- Montgomery, R.H., J.C. Loftis and J. Harris. 1987. Statistical characteristics of ground-water quality variables. *Ground Water*. 25(2):176-184.
- Montgomery, R.H. and J.C. Loftis. 1988. Reply to discussion by Dennis R. Helsel and Robert M. Hirsch "Applicability of the t-test for detection trends in water quality variables." *Water Resources Bulletin*. 24(1):205-207.
- Natrella, M.G. 1960. The relation between confidence intervals and tests of significance. *Statistical Issues—A Reader for the Behavioral Sciences*. R.E. Kirk, ed. Brooks/Cole Publishing Company, Monterey, CA, 1972. pp. 113-117.
- Nelson, D.W. and R.H. Dowdy (co-chairs of workshop and editors). 1988. *Methods for ground water quality studies*. Proceedings of a national workshop held at Arlington, Virginia on November 1-3, 1988. Published by Agricultural Research Division, University of Nebraska-Lincoln, Lincoln, NE.
- Newell, C.J., L.P. Hopkins and P.B. Bedient. 1990. A hydrogeologic database for ground-water modeling. *Ground Water*. 28(5):703-714.
- Newman, M.C., P.M. Dixon, B.B. Looney and J.E. Pinder III. 1989. Estimating mean and variance for environmental samples with below detection limit observations. *Water Resources Bulletin*. 25(4):905-916.
- NRC. 1990. *Managing troubled waters—the role of marine environmental monitoring*. A report by the Committee on a Systems Assessment of Marine Environmental Monitoring, Marine Board, Commission on Engineering and Technical Systems, National Research Council. National Academy Press, Washington D.C.
- O'Brien, T.C. and B.J. Shapiro. 1968. *Statistical significance—what?* *Statistical Issues—A Reader for the Behavioral Sciences*. Brooks/Cole Publishing Company, Monterey, CA, 1972.
- Ozonoff, D. and L.I. Boden. 1987. Truth and consequences: health agency responses to environmental health problems. *Science, Technology and Human Values*. 12(3&4):70-77.
- Patel, H.I. and G.D. Gupta. 1984. A problem of equivalence in clinical trials. *Biometrical Journal*. 26(5):471-474.

- Pavur, R. 1988. Type I error rates for multiple comparison procedures with dependent data. *The American Statistician*. 42(3):171-173.
- Perry, J.N. 1986. Multiple-comparison procedures: a dissenting view. *FORUM: Journal of Economic Entomology*. 79(5):1149-1155.
- Phillips, R.D, J.C. Loftis and H.P. Hotto. 1989. User's manual for WQSTAT II—a water quality statistics program. Colorado State University, Fort Collins, CO.
- Porter, P.S. 1986. Statistical analysis of water quality data affected by the limits of detection. Unpublished Ph.D. Dissertation, Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, CO.
- Porter, P.S., R.C. Ward and H.F. Bell. 1988. The detection limit. *Environmental Science and Technology*. 22(8):856-861.
- Porter, P.S. and R.C. Ward. 1991. Estimating central tendency from uncensored trace level measurements. *Water Resources Bulletin*. 27(4):687-700.
- Quenouille, M.H. 1949. Approximate tests of correlation in time series. *Journal of the Royal Statistical Society series B*. 11:68-84.
- Reed, M. and V.J. Bierman, Jr. 1983. Proceedings of a Workshop for the Development of a Scientific Protocol for Ocean Dumpsite Designation. Held at the W. Alton Jones Campus of the University of Rhode Island, February 14-17, Kingston, Rhode Island. U.S. Environmental Protection Agency Report No. ASA/RI-83-001.
- Richards, R.P. 1989. Evaluation of some approaches to estimating non-point pollutant loads for unmonitored areas. *Water Resources Bulletin*. 25(4):891-904.
- Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian and V. Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, CO.
- Sara, M.N. and R. Gibbons. 1991. Organization and analysis of water quality data. *Practical Handbook of Ground-Water Monitoring*. D.M. Neilson ed. Lewis publishers, Chelsea, MI. pp. 541-588.
- Scariano, S.M. and J.M. Davenport. 1987. The effects of violations of independence assumptions in the one-way ANOVA. *The American Statistician*. 41(2):123-129.
- Schaub, S.A. and C.P. Gerba. 1988. Guide standard and protocol for testing microbiological water purifiers. Proceedings—Conference on Point-of-Use Treatment of Drinking Water. Cincinnati, Ohio, October 6-8, 1987. U.S. Environmental Protection Agency Report No. EPA/600/9-88/012.
- Schilling, E.G. and P.R. Nelson. 1976. The effect of non-normality on the control limits of X charts. *Journal of Quality Technology*. 8(4):183-188.
- Schmid, C.F. 1983. *Statistical graphics—design principles and practices*. John Wiley and Sons, New York, NY.
- Schroder, L.J. and H.E. Taylor. 1980. Contamination studies and testing protocol for polyethylene bottles. U.S. Geological Survey Open File Report 80-165.

References

- Schubel, J.R. 1987. Program accountability—how do we measure progress?—A scientific view. Estuarine and coastal management—tools of the trade. Proceedings of the Tenth National Conference of The Coastal Society. October 12-15, 1986. New Orleans, LA. Copyright by The Coastal Society, 1987.
- Schuk, W.W. 1986. Evaluation of a dissolved oxygen field test protocol. Aeration systems—design, testing, operation, and control. W.C. Boyle, ed. Noyes Publications, Park Ridge, NJ.
- Sen, P.K. and M.L. Puri. 1977. Asymptotically distribution-free aligned rank order tests for composite hypotheses for general multivariate linear models. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 39:187-195.
- Smith, D.G. and G.B. McBride. 1990. New Zealand's national water quality monitoring network—design and first year's operation. *Water Resources Bulletin.* 26(5):767-775.
- Snedecor, G.W. and Cochran, W.G. 1989. *Statistical Methods.* 8th Ed. Iowa State University Press, Ames, IA.
- Starks, T.H. and G.T. Flatman. 1991. RCRA ground-water monitoring decision procedures viewed as quality control schemes. *Environmental Monitoring and Assessment.* 16:19-37.
- Statgraphics®. 1991. Reference manual—Statgraphics® V.5. STSC, Rockville, MD.
- Steen, W.C. 1988. Project summary—interim protocol for measuring microbial transformation rate constants for suspended bacterial populations in aquatic systems. U.S. Environmental Protection Agency Report No. EPA/600/S3-88/007.
- Taylor, J.K. 1987. *Quality Assurance of Chemical Measurements.* Lewis Publishers, Chelsea, MI.
- Thomas, R.L. 1987. A protocol for the selection of process-oriented remedial options to control in situ sediment contaminants. *Hydrobiologia.* 149:247-258.
- Thompson, C.R., G. Kats, P. Dawson and D. Doyle. 1981. Project summary—development of a protocol for testing the effects of toxic substances on plants. U.S. Environmental Protection Agency Report No. EPA-600/S3-81-006.
- Tondeur, Y., W.F. Beckert, S. Billets and R.K. Mitchum. 1989. Method 8290: an analytical protocol for the multimedia characterization of polychlorinated dibenzodioxins and dibenzofurans by high-resolution gas chromatography/high-resolution mass spectrometry. *Chemosphere.* 18:119-131.
- Tufte, E.R. 1983. *The Visual Display of Quantitative Information.* Graphic Press, Cheshire, CT.
- Tufte, E.R. 1990. *Envisioning Information.* Graphics Press, Cheshire, CT.
- Tukey, J.W. 1977. *Exploratory Data Analysis.* Addison-Wesley Publishing Company, Reading, MA.
- U.S. Department of Labor. 1967. The consumer price index: history and techniques. U.S. Department of Labor, Bureau of Labor Statistics, Bulletin No. 1517.
- U.S. EPA. 1985. Resource Document for the Ground-Water Monitoring Strategy Workshop. Office of Ground-Water Protection, U.S. Environmental Protection Agency, Washington D.C.

- U.S. EPA. 1986. RCRA ground water monitoring technical enforcement guidance document (final). Office of Waste Programs Enforcement, U.S. Environmental Protection Agency, Washington, D.C. OSWER Report No. 9950.1.
- U.S. EPA. 1988. Guidance on Remedial Actions for Contaminated Ground Water at Superfund Sites. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington D.C.
- U.S. EPA. 1989. Statistical analysis of ground-water monitoring data at RCRA facilities—interim final guidance. Office of Solid Waste, Waste Management Division, U.S. Environmental Protection Agency, Washington, D.C.
- U.S. EPA. 1990. The Nation's Hazardous Waste Management Program at a Crossroads: The RCRA Implementation Study. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington D.C.
- Vaughan, W.J. and C.S. Russell. 1983. Monitoring point sources of pollution: answers and more questions from statistical quality control. *The American Statistician*. 37(4):476-487.
- Velleman, P.F. and Hoaglin, D.C. 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston, MA.
- Wainer, H. and D. Thissen. 1981. Graphical data analysis. *Annual Review of Psychology*. 32:191-241.
- Walker, W.W. 1991. Water quality trends at inflows to Everglades National park. *Water Resources Bulletin*. 27(1):59-72.
- Wallis, J.R. and P.E. O'Connell. 1972. Small sample estimation of r_1 . *Water Resources Research*. 8(3):707-212.
- Ward, R.C., J.C. Loftis, H.P. DeLong and H.F. Bell. 1988. Groundwater quality: a data analysis protocol. *Journal Water Pollution Control Federation*. 60(11):1938-1945.
- Ward, R.C. and J.C. Loftis. 1989. Monitoring systems for water quality. *Critical Reviews in Environmental Control*. 19(2):101-118.
- Ward, R.C., J.C. Loftis and G.B. McBride. 1990. Design of Water Quality Monitoring Systems. Van Nostrand Reinhold Company, NY.
- Weinberg, S. and K. Goldberg. 1990. Statistics for the Behavioral Sciences. Cambridge University Press, Cambridge, England.
- Wilson, A.L. 1973. The performance characteristics of analytical methods—III. *Talanta*. 20:725-732.
- Wolbach, C.D., R.R. Whitney and U.B. Spannagel. 1984. Project summary—design and development of a hazardous waste reactivity testing protocol. U.S. Environmental Protection Agency Report No. EPA-600/S2-84-057.
- Wurman, R.S. 1989. Information Anxiety. Doubleday, New York, NY.
- Zahn, D.A. and Isenberg, D.J. 1983. Nonstatistical aspects of statistical consulting. *The American Statistician*. 37(4), November.
- Zar, J.H. 1984. Biostatistical Analysis. 2nd Ed. Prentice-Hall, Englewood Cliffs, NJ.

Data Analysis Protocol
for Ground Water Quality Monitoring:
IBM East Fishkill Facility

Table of Contents

Glossary of Acronyms	vi
I. Introduction	1
A. Purpose	1
B. Scope.....	1
C. Organization.....	1
II. Information Goals	1
A. Regulatory Information Goals	1
B. Monitoring Information Goals	3
C. Statistical Information Goals.....	3
III. Data Record Attributes	3
A. General	3
B. Multiple Observations	3
C. Outliers.....	3
D. Changing Sampling Frequencies.....	4
E. Missing Values	4
F. Nonnormality	4
G. Seasonality	4
H. Censoring.....	4
I. Serial Correlation.....	4
IV. Data Analysis Methods	4
A. General	4
B. Computational Methods.....	5
C. Graphical Methods	5
D. Direct Comparison with Fixed Limits	5
E. Statistical Methods	5
1. Control Charts	5
2. Interval Estimation of the Difference in Medians.....	7
3. Interval Estimation of the Median	7

V. Criteria for Choosing Compounds for Laboratory Analysis, Data Summarization and Data Analysis	8
A. Routinely Monitored Compounds	8
B. Selected Compounds	8
C. Key Compounds	8
VI. Criteria for Choosing Sampling Intervals	8
VII. Procedures for Reviewing the Data Analysis Protocol	8
VIII. References	9
Appendix A—Data Analysis Protocol Summary Sheets	11
Appendix B—Sampling Frequencies	25
Appendix C—Categories of Compounds	35
Appendix D—Reporting Formats	41
Appendix E—Data Analysis Methods	45
Appendix E-1. Combined Shewart-CUSUM control charts.	46
Appendix E-2. The Hodges-Lehmann estimator and confidence interval..	55
Appendix E-3. Interval estimate of the median	58

List of Tables

1. Information goals and fixed limits for direct comparison	6
B-1. Sampling frequencies (per year)—area A remediation area.....	26
B-2. Sampling frequencies (per year)—area B remediation area.	26
B-3. Sampling frequencies (per year)—area C remediation area.	27
B-4. Sampling frequencies (per year)—area D remediation area.	27
B-5. Sampling frequencies (per year)—SEQ remediation area.....	28
B-6. Sampling frequencies (per year)—area C B/330 investigative area.....	29
B-7. Sampling frequencies (per year)—building 322 investigative area.....	30
B-8. Sampling frequencies (per year)—bedrock aquifer.....	31
B-9. Sampling frequencies (per year)—west complex.....	33
C-1. "Appendix 33" compounds.....	36
C-2. Routinely monitored compounds.....	36
C-3. Selected compounds.....	38
C-4. Key compounds.....	39

List of Figures

1. Laws and regulations governing ground water quality monitoring at the IBM East Fishkill facility	2
D-1. Reporting format for combined Shewart-CUSUM control chart	42
D-2. Reporting format for Hodges-Lehmann estimates and confidence intervals...	42
D-3. Reporting format for time series plots.	43
D-4. Reporting format for interval estimates of the median.	43
E-1. Worksheet for designing combined Shewart-CUSUM control charts.....	49
E-2. Design of combined Shewart-CUSUM control chart for information goal #2.....	50
E-3. Worksheet for implementing combined Shewart-CUSUM control charts	51
E-4. Implementation of combined Shewart-CUSUM control chart for information goal #2	52
E-5. Plot of Shewart-CUSUM data from Figure E-4.....	53
E-6. Comparison of confidence intervals to the ground water protection standard (GWPS).....	60

Glossary of Acronyms

AOC	area of concern
ASTM	American Society for Testing and Materials
CSU	Colorado State University
CUSUM	cumulative sum
DAP	data analysis protocol
GC/MS	gas chromatograph/mass spectrometer
GWPS	ground water protection standard
MDL	method detection limit
NYS	New York State
NYSDEC	New York State Department of Environmental Conservation
RCRA	Resource Conservation and Recovery Act
SEQ	southeast quadrant

I. INTRODUCTION

A. Purpose

A protocol for using data analysis methods to obtain information from IBM's ground water monitoring program is presented in this document. The data analysis protocol (DAP) will help to ensure that all collected data has a specified purpose, and that statistics is properly used in both a theoretical and practical sense.

It is anticipated that the DAP will provide numerous benefits including:

1. Those involved in data collection will develop a better understanding of the importance of producing high quality data.
2. The subjectivity of statistical analysis will be attenuated because statistical methods will be specified prior to data collection.
3. Economic benefits will be realized because only data which contributes to useful information will be collected.
4. Continuity of data analysis in the event of employee turnover will be facilitated.
5. Data analysis methods will be correctly applied, resulting in reliable information.
6. Because they are well documented, data analysis procedures will be auditable.
7. Individuals who are involved with management and operation of the monitoring program will have a basis for communication.
8. Water quality information will be extracted from the data as soon as possible.
9. Monitoring system operation will be driven by information goals rather than by politics or short term crises.

The DAP was written by Colorado State University researchers and IBM ground water staff using a framework developed by Adkins (1992). The protocol represents IBM's philosophy to "meet or exceed all regulatory requirements." It also meets RCRA goals of being protective of human health and the environment.

B. Scope

The original intent of the DAP was to provide guidelines for using statistical methods to

analyze ground water quality data. Soon after development of the protocol started, however, it became apparent that the scope should be expanded to include non-statistical methods. It also became evident that certain aspects of network design, such as sampling frequencies and compounds to be analyzed, should be included in the protocol.

The primary intended users of the DAP are ground water professionals at the IBM East Fishkill Facility. General managers, field samplers and laboratory personnel should also find the protocol useful.

The DAP addresses monitoring primarily from a long-term management point of view. DAPs for investigative studies will be written separately.

C. Organization

The majority of detailed information in the protocol is presented as appendices. Appendix A consists of DAP summary sheets which comprise the supporting structure for the rest of the protocol. Protocol organization is clearly shown in the table of contents.

II. INFORMATION GOALS

A. Regulatory Information Goals

Regulatory information goals were identified by reviewing legislation and by conversing with regulators. An outline of applicable laws and regulations is presented in Figure 1. The regulatory information goals which were identified are as follows:

- Preserve and enhance the quality of water resources.
- Protect human health and the environment.
- Determine background water quality.
- Characterize the extent, composition, concentration and movement of contaminant releases.
- Prevent contaminants from migrating off-site.
- Demonstrate the effectiveness of corrective action.

U.S. Code

Title 42: The Public Health and Welfare

Chapter 82: Solid Waste Disposal

Subchapter III: Hazardous Waste Management

Section 6924: Standards applicable to owners and operators of hazardous waste treatment, storage, and disposal facilities

(p) Ground water monitoring

Section 6926: Authorized State hazardous waste programs

(b) Authorization of State programs

The State of New York was given authorization to manage their own hazardous waste program on May 15, 1986 (51 Federal Register 17737).

Code of Federal Regulations

Title 40: Protection of Environment

Chapter 1: Environmental Protection Agency

Subchapter I: Solid Wastes

Part 264: Standards for owners and operators of hazardous waste treatment, storage, and disposal facilities

Part 271: Requirements for authorization of State hazardous waste programs

New York Environmental Conservation Law

Article 27: Waste and Refuse

Title 9: Industrial Hazardous Waste Management

Section 27-0911: Standards applicable to owners and operators of hazardous waste treatment, storage, and disposal facilities

New York Code of Rules and Regulations

Title 6: Environmental Conservation

Subpart 373-2: Final status standards for owners and operators of hazardous waste treatment, storage, and disposal facilities

Section 373-2.6: Releases from solid waste management units

(l) Corrective action for solid waste management units

New York Permit Writers' Guide

373 Generic (5/14/91)

Module III: Corrective Action Requirements for Solid Waste Management Units and Areas of Concern

E.: Corrective Action Requirements

Figure 1. Laws and regulations governing ground water quality monitoring at the IBM East Fishkill facility.

B. Monitoring Information Goals

Monitoring information goals which were identified for the data analysis protocol are listed here. Summary sheets for each information goal are presented in Appendix A.

- #1. Ensure adequacy of the treatment facility by screening for the presence of new compounds.
- #2. Ensure adequacy of the treatment facility by detecting changes in input concentrations.
- #3. Determine the lateral extent of contamination plumes.
- #4. Determine the mass of contaminants removed at each remediation area.
- #5. Characterize contaminant concentrations over time within the plumes.
- #6. Determine the hydraulic effectiveness of remediation.
- #7. Measure water quality down-gradient of AOCs where flow may pass across a site boundary.
- #8. Confirm previous monitoring results which indicate that contaminants are not migrating offsite from onsite.
- #9. Confirm previous monitoring results which indicate that contaminants are not migrating onsite from offsite.
- #10. Ensure that all chemicals of concern are properly identified and monitored.
- #11. Provide a qualitative basis for comparison with data from "Appendix 33" monitoring of contaminated wells.
- #12. Provide data for investigative studies associated with the NYS Part 373 Corrective Action Program.

C. Statistical Information Goals

Statistical methods were chosen to analyze data for three of the monitoring information

goals. Those methods are described in detail in Appendix E. Statistical information goals are restated here for monitoring goals #2, #3 and #7 respectively.

- Immediately detect concentrations which are $\geq 4\sigma$ above the target mean. Quickly detect persistent changes in concentration which are $\geq 1.5\sigma$ above the target mean.
- Estimate the magnitude of the difference between the current two years of concentration data and the previous two years of concentration data. Determine a 90 percent confidence interval for the estimate.
- On an annual basis, estimate the 95 percent confidence interval of the true population median concentration.

III. DATA RECORD ATTRIBUTES

A. General

Data record attributes are characteristics of data which can complicate statistical analysis (Bell and Delong, 1988). The following data record attributes are addressed in this protocol:

- multiple observations
- outliers
- changing sampling frequencies
- missing values
- nonnormality
- seasonality
- censoring
- serial correlation

B. Multiple Observations

Multiple observations occur when replicate samples are collected for QA/QC purposes. If a single value is needed for statistical analysis, the multiple values will be averaged.

C. Outliers

Outliers are values which are obviously higher or lower than the majority of the data set. Unless there is evidence to show that they are erroneous, outliers will be retained and analyzed with the rest of the data.

D. Changing Sampling Frequencies

Most statistical methods which examine the behavior of water quality data over time require equally spaced samples. Consistency of sampling intervals with past data was considered when frequencies were chosen for the DAP. Future information needs were also appraised. Although some changing sampling frequencies are anticipated at the IBM site, they are not expected to affect any of the data analysis procedures recommended in the DAP.

E. Missing Values

Some statistical analysis techniques which require equal sample sizes or regularly spaced samples in time, cannot be applied to data records which have missing values. A great deal of effort is made at the East Fishkill facility to avoid missing values because of the interference they cause with data analysis methods. It is IBM's policy to not replace missing values with any type of numerical response. If missing values appear to have a significant impact on a particular data analysis method, an appropriate note will accompany the results.

F. Nonnormality

Nonnormality is a common characteristic of water quality variables. Water quality data is often right-skewed because it has a lower bound of zero, many values near the detection limit, and infrequent high values.

Characterization studies conducted on ground water quality data from the IBM East Fishkill site indicate that the data has slight to moderate nonnormality. Specific information on how nonnormality will be handled when statistically analyzing ground water quality data is presented in section IV.E.

G. Seasonality

Seasonality is the change in distribution of water quality variables which can be attributed to the time of year. A "season" can be any specified period of time but is generally one month (twelve seasons per year) or three months (four seasons per year). Seasonality

may or may not occur as a consistent pattern. Characterization studies conducted on ground water quality data from the IBM East Fishkill site indicate that moderate seasonality is present in data collected from the shallow aquifer. In general, IBM has dealt with seasonality according to recommendations provided in Ward et al. (1988). Specific information on how seasonality will be handled when statistically analyzing ground water quality data is presented in section IV.E.

H. Censoring

IBM has not censored laboratory data since 1986. This noncensoring approach is recommended by Porter (1986), Porter et al. (1988), Ward et al. (1988), and ASTM Method D4210-83. IBM's laboratory reports the actual readings together with an estimate of measurement error. The uncensored laboratory measurements are used for all data calculations, and the results are rounded to the appropriate number of significant figures prior to reporting.

I. Serial correlation

Serial correlation is generally thought of as redundancy of information between adjacent observations in a time series. Serial correlation causes data to violate the assumption of independence which underlies most statistical methods.

Characterization studies conducted on ground water quality data from the IBM East Fishkill site indicate that serial correlation is minimal for quarterly sampling frequencies. Specific information on how serial correlation will be handled is presented in section IV.E.

IV. DATA ANALYSIS METHODS

A. General

Several questions dealing with information goals, data characteristics, and data analysis method characteristics were addressed during the process of choosing data analysis methods. Some of the questions are listed here:

Information Goals

- What do we want to know?
- How soon do we need to know it?
- What degree of accuracy is appropriate?

Data Record Attributes

- What can historical data tell us about data record attributes?
- How should we deal with data record attributes?
- How can we prevent missing values and changing sampling frequencies?

Characteristics of the Data Analysis Method

- Is the data analysis method easy to understand?
- Is it widely accepted?
- Does it tell us what we need to know?

Data analysis methods which were chosen are discussed below. Advantages and limitations of the techniques are summarized, and the rationale used to choose them is described. Technical details, such as calculation procedures, are presented in Appendix E. For purposes of the following discussion, the methods are divided into four categories:

- Computational.
- Graphical.
- Direct comparison with fixed limits.
- Statistical.

B. Computational Methods

Strictly computational procedures (i.e., those involving calculations but not statistics) were chosen for information goal #4: *determine the mass of contaminants removed at each remediation area.*

C. Graphical Methods

Plots of concentration versus time were selected to analyze data for information goal #5: *characterize contaminant concentrations over time within the plumes.* This qualitative, graphical approach was chosen because it provides information on general patterns of contaminant behavior. Quantitative trend analysis, such as estimation of trend magnitude or determination of statistical significance, is not needed at this time. It may, however, be appro-

priate in the future when data records are longer, and when the relationship between remediation effectiveness and trend is better understood. Current data collection practices—fixed interval sampling, no censoring and virtually no missing values—produce the type of data which is needed for quantitative trend analysis should the need arise.

Data collected for information goal #6: *determine the hydraulic effectiveness of remediation,* will be converted into information by plotting the water levels on a site map and constructing contours of the water table.

D. Direct Comparison With Fixed Limits

Direct comparison with fixed limits was selected as a technique to analyze data for five of the twelve information goals. The method involves comparison of individual concentration values to fixed limits without the use of statistics.

The five information goals for which direct comparison will be used are listed in Table 1. Fixed limits which have been set for each goal are also listed.

The method of direct comparison with fixed limits was chosen for the information goals listed in Table 1 because it was felt that action should be taken if a single verified value falls above the specified limits. Recommended action for each information goal is described in Appendix A.

E. Statistical Methods

E.1 Control Charts

Control chart methods were chosen to analyze data for information goal #2: *ensure adequacy of the treatment facility by detecting changes in input concentrations.*

Control charts are graphical representations of statistical quality control procedures. They consist of a horizontal line corresponding to the average value of the characteristic in question, together with upper and/or lower control limits (Marriott, 1990). Most commonly, the horizontal axis is in units of time and the vertical axis is in units of standard deviation.

Table 1. Information goals and fixed limits for direct comparison.

	Information Goal	Fixed Limits
#1:	Ensure adequacy of the treatment facility by screening for the presence of new compounds.	MDL
#7:	Measure water quality down-gradient of AOC's where flow may pass across a site boundary.*	2 x MDL
#8:	Confirm previous monitoring results which indicate that contaminants are not migrating offsite from onsite.	GWPS
#9:	Confirm previous monitoring results which indicate that contaminants are not migrating onsite from offsite.	GWPS
#10:	Ensure that all chemicals of concern are properly identified and monitored.	MDL

*Interval estimation is the primary procedure chosen for information goal #7 (see the "statistical methods" section). Comparison with fixed limits was chosen as a secondary method to ensure that very high values are dealt with immediately.

Sampling results or statistics of sampling results are plotted sequentially on the chart. If a plotted point falls outside the limits, the process is said to be out of control.

"Out of control" simply means that statistically unusual variation from normal performance has been detected (Berthouex et al., 1978). Action taken in response to an out of control signal may consist of increasing the retention time in a clarifier, adjusting manufacturing equipment, or just taking a closer look at the data.

Control charts have traditionally been used for process quality control in the manufacturing industry. In the field of water quality management, control charts are used to monitor sewage treatment plant effluent (Berthouex and Hunter, 1975; Berthouex et al., 1978; and Vaughan and Russell, 1983) and to analyze water quality data at RCRA sites (U.S. EPA, 1989 and Starks and Flatman, 1991).

Environmental quality data often violates the underlying assumptions of control charts by exhibiting serial correlation, seasonality,

nonconstant variance and nonnormal distributions (Berthouex et al., 1978). Complications arising from "messy" data are exacerbated by the small sample sizes encountered in ground water quality monitoring. Research which addresses these issues is still in its early stages. The results of control chart analysis of ground water quality data are therefore subject to a high degree of uncertainty and should be used accordingly. It would be inappropriate to base major environmental decisions on the results of control chart analyses.

If their limitations are acknowledged, however, control charts can be an effective tool for quickly detecting shifts in water quality. They are versatile and easy to use. Most importantly, they provide an ongoing visual account of water quality. Research has shown that shifts in quality are often more readily apparent on CUSUM control charts than on time series plots (Hockman and Lucas, 1987).

Results of control chart analyses will be used as a tool to optimize the efficiency of IBM's ground water treatment facilities. Results will

not be used as the basis for regulatory action.

The decision was made to employ a control chart procedure known as the "combined Shewart-CUSUM method" (Lucas, 1982) to analyze data for information goal #2. Large shifts in concentration values are detected by Shewart limits, whereas small changes which persist are detected by CUSUM limits. The combined Shewart-CUSUM method is easier to apply than other control chart methods which provide similar results (Lucas, 1982). Details of the procedure are presented in Appendix E-1.

Evaluation of existing data indicates that low to moderate amounts of serial correlation, seasonality and nonnormality should be expected in the data which is to be analyzed by control chart methods. Serial correlation and seasonality increase the probability of Type I error (i.e., false positives) (Loftis et al., 1987 and U.S. EPA, 1989). The effects of nonnormality depend on the actual distribution which is present (Schilling and Nelson, 1976).

No special adjustments will be made to the combined Shewart-CUSUM method to account for serial correlation, seasonality or nonnormality. This decision was based on three factors:

1. Adjustments may complicate the method thereby making it more difficult to understand from an intuitive point of view.
2. Modifications which involve data transformations can decrease information content of the data and/or make visual inspection of the data less effective.
3. Modifications to account for data record attributes may affect the results of control charts in ways which are not well understood.

E.2 Interval Estimation of the Difference in Medians

Interval estimation was one of the methods chosen to *determine the lateral extent of contamination plumes* (information goal #3). An interval estimate of the difference between medians for the current two years of data and the previous

two years of data will be made for each key compound (or for each sum of key compounds). The estimates will indicate whether or not concentrations have changed enough in the well to warrant relocation of the plume boundary. (A more detailed explanation of how plume boundaries will be drawn is given in the summary sheet for information goal #3.)

An interval estimate is an intuitively appealing way of presenting information. The width of the confidence interval indicates how much reliance should be placed on the estimate. Similar information can be obtained from the yes/no results of a hypothesis test accompanied by the appropriate operating characteristic curve (Natrella, 1972). Operating characteristics curves, however, are difficult to construct and interpret.

The particular estimator chosen for information goal #3 is the Hodges-Lehmann estimator (Hodges and Lehmann, 1963). It is simply the median of all pairwise differences between the two groups. The Hodges-Lehmann estimator was chosen because it is robust in the presence of serial correlation, seasonality and nonnormality, all of which may be present in small to moderate amounts (Hirsch, 1988). Details of the procedure are given in Appendix E-2.

E.3 Interval Estimation of the Median

Interval estimation of the median was the method selected to analyze data for information goal #7: *measure water quality down-gradient of AOCs where flow may pass across a site boundary*. Interval estimates of the median will be made on a yearly basis and compared to the ground water protection standard (GWPS). Action will be taken if the entire interval is above the GWPS.

A statistical approach was chosen instead of direct comparison because some of the compounds are expected to be present at levels above the MDL. Direct comparison is appropriate when the main objective is to detect compounds which have not previously been found in a particular well. The objective here, however, is to determine if any compounds

are present at levels above the GWPS, while accounting for distributional variability.

Details of the interval estimation procedure are presented in Appendix E-3.

V. CRITERIA FOR CHOOSING COMPOUNDS FOR LABORATORY ANALYSIS, DATA SUMMARIZATION AND DATA ANALYSIS

A. Routinely Monitored Compounds

All samples will be analyzed for the "routinely monitored compounds" which are listed in Table C-2. Results will be reported as raw data.

B. Selected Compounds

A procedure which employed boxplots was used to decide which of the routinely monitored compounds should be designated as "selected compounds." Boxplots were constructed for data from the second quarter of 1989 through the first quarter of 1991. The data represented either single compounds or groups of compounds. Compounds (or groups of compounds) whose upper fence fell above the ground water protection standard (usually $5\mu/L$) were designated as selected compounds.

Selected compounds are listed in Table C-3. Results of analyses of selected compounds will be reported as summary statistics.

C. Key Compounds

Key compounds are those routinely monitored compounds which are examined by special data analysis techniques. Key compounds vary according to area as shown in Table C-4. They were chosen based on concentration, importance, and their ability to meet information goals.

VI. CRITERIA FOR CHOOSING SAMPLING INTERVALS

Factors which were considered when choosing sampling intervals for the data analysis protocol included:

- Regulatory requirements and guidance, which generally recommended a minimum of semi-annual sampling.
- Monitoring and statistical information goals.
- Historical water quality records for the well and adjacent points.
- Location of the well and direction of ground water flow.
- Risk to human health and the environment.

VII. PROCEDURES FOR REVIEWING THE DATA ANALYSIS PROTOCOL

The data analysis protocol will be reviewed once every two years. It will be reviewed more frequently if any of the following occur:

- New regulations are promulgated which affect the content of the DAP.
- It becomes evident that a section of the DAP is invalid.
- A particular revision would greatly improve the effectiveness of the DAP.

Minor revisions will be saved for the next scheduled review.

Protocol review will be overseen by the ground water quality program manager. The manager will sign-off on all revisions.

Factors which will be considered during the reviews include current research, amended regulations, and new information obtained from recent monitoring data.

All revisions made to the data analysis protocol will be thoroughly documented and, if necessary, reported to NYSDEC. Each revision will have the following heading:

GROUND WATER QUALITY DATA ANALYSIS PROTOCOL REVISION

Date:

Author:

Approved by:

Summary of Revision:

All copies of the DAP will be kept in loose-leaf notebooks to facilitate the revision process. Revision sheets will be placed at the back of the protocol. Sections of the protocol

which are no longer applicable will be crossed out but not removed.

With a loose-leaf format, the potential exists for sheets to be lost or disorganized. To remedy

this situation, one copy of the DAP will be designated as the master copy. It will be carefully maintained and kept in one location.

VIII. REFERENCES

- Adkins, N.C. 1992. A framework for development of data analysis protocols for groundwater quality monitoring systems. Unpublished Ph.D. dissertation, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.
- Bell, H.F. and H.P. DeLong. 1988. Data characteristics: groundwater monitoring's "Catch 22." American Chemical Society—75 Years of Progress. Preprints of papers presented at the 196th ACS national meeting, Los Angeles, CA, September 25039, 1988. 28(2):20-24,
- Berthouex, P.M. and W.G. Hunter. 1975. Treatment plant monitoring programs: a preliminary analysis. *Journal Water Pollution Control Federation*. 47(8):2143-2156.
- Berthouex, P.M., W.G. Hunter and L. Pallesen. 1978. Monitoring sewage treatment plants: some quality control aspects. *Journal of Quality Technology*. 10(4):139-149.
- Bissell, A.F. 1984. The performance of control charts and Cusums under linear trend. *Applied Statistics*. 33(2):145-151.
- Cheremisinoff, N.P. 1988. Computer statistics for environmental engineers. *Pollution Engineering*. 20(1):54-62.
- Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. Elsevier Publishers.
- Hirsch, R.M. 1988. Statistical methods and sampling design for estimating step trends in surface-water quality. *Water Resources Bulletin*. 24(3):493-503.
- Hockman, K.K. and J.M. Lucas. 1987. Variability reduction through subvessel CUSUM control. *Journal of Quality Technology*. 19(3):113-121.
- Hodges, J.L. and E.L. Lehmann. 1963. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*. 34:598-611.
- Hollander, M. and D.A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley and Sons, New York, NY.
- Loftis, J.C., P.S. Porter and G. Settembre. 1987. Statistical analysis of industrial wastewater monitoring data. *Journal Water Pollution Control Federation*. 59(3):145-151.
- Lucas, J.M. 1982. Combined Shewart-CUSUM quality control schemes. *Journal of Quality Technology*. 14(2):51-59.
- Lucas, J.M. and R.B. Crosier. 1982. Robust CUSUM: a robustness study for CUSUM quality control schemes. *Communications in Statistics—Theory and Methods*. 11(23):2669-2687.

- Lucas, J.M. 1985. Cumulative Sum (CUSUM) control schemes. *Communication in Statistics—Theory and Methods*. 14(11):2689-2704.
- Marriott, F.H.C. 1990. *A Dictionary of Statistical Terms*. Longman Scientific and Technical, Essex, England, and John Wiley and Sons, Inc., New York, NY.
- Natrella, M.G. 1972. The relationship between confidence intervals and tests of significance. *Statistical Issues: A Reader for the Behavioral Sciences*. R.E. Kirk ed., Brooks/Cole Publishing Co., Monterey, California. p.113-117.
- Page, E.S. 1954. Continuous inspection schemes. *Biometrika*. 41(1,2):100-114.
- Page, E.S. 1961. Cumulative sum charts. *Technometrics*. 3(1):1-9.
- Porter, P.S. 1986. Statistical analysis of water quality data affected by the limits of detection. Unpublished Ph.D. Thesis, Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, Colorado.
- Porter, P.S., R.C. Ward and H.F. Bell. 1988. The detection limit. *Environmental Science and Technology*. 22(8):856-861.
- Schilling, E.G. and P.R. Nelson. 1976. The effect of non-normality on the control limits of X charts. *Journal of Quality Technology*. 8(4):183-188.
- Starks, T.H. 1989. Evaluation of control chart methodologies for RCRA (Resource Conservation and Recovery Act) Waste Sites. Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, Nevada. U.S. EPA Report No. EPA/600/4-88/040.
- Starks, T.H. and G.T. Flatman. 1991. RCRA ground-water monitoring decision procedures viewed as quality control schemes. *Environmental Monitoring and Assessment*. 16:19-37.
- U.S. EPA. 1989. Statistical analysis of ground-water monitoring data at RCRA facilities—interim final guidance. Office of Solid Waste, Waste Management Division, U.S. Environmental Protection Agency, Washington, D.C., February.
- Vasilopoulos, A.V. and A.P. Stamboulis. 1978. Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology*. 10(1):20-29.
- Vaughan, W.J. and C.S. Russell. 1983. Monitoring point sources of pollution: answers and more questions from statistical quality control. *The American Statistician*. 37(4):476-487.
- Ward, R.C., J.C. Loftis, H.P. DeLong and H.F. Bell. 1988. Groundwater quality: a data analysis protocol. *Journal Water Pollution Control Federation*. 60(11):1938-1945.

Appendix A

Data Analysis Protocol Summary Sheets

INFORMATION GOAL #1: Ensure adequacy of the treatment facility by screening for the presence of new¹ compounds.

Compounds To Be Analyzed: "Appendix 33" compounds (Table C-1).

Well Type: Extraction wells

Number of Wells To Be Sampled: 7

Sampling Frequency: 1 per year for 3 years

Data Analysis Procedures:

Concentrations will be compared with the method detection limit (MDL).

Actions To Be Taken Based On Laboratory Results:

If a new compound is detected at or above the MDL, the well will be resampled. If the concentration is still at or above the MDL, then the presence of the compound is confirmed², and:

- (1) The pertinent well, compound and concentration will be reported to the New York State Department of Environmental Conservation (NYSDEC) within seven days of confirmation.
- (2) If necessary, adjustments will be made to the treatment system to account for the presence of the new compound.
- (3) If the new compound is not already included in the list of routinely monitored compounds (Table C-2), a decision will be made regarding whether or not to add it to the list. The decision will be based on several factors including the concentration of the compound in relation to the GWPS, and the location and direction of ground water flow.

Reporting:

Results will be presented in the first semi-annual report which is written subsequent to sampling as follows:

- (1) Results of all laboratory analyses will be reported in a raw data format.
- (2) If a resample is conducted and the compound of interest is no longer at or above the MDL, then the original analysis will be labeled "invalid."
- (3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

¹ "New" compounds are those which have not been previously detected and confirmed in a particular well.

² Organics will be analyzed by GC/MS where applicable in order to yield positive qualitative identification.

INFORMATION GOAL #2: Ensure adequacy of the treatment facility by detecting changes in input concentrations.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type: Extraction wells

Number Of Wells To Be Sampled: 12

Sampling Frequency: 12 per year

Data Analysis Procedures:

All key compounds (Table C-4) will be tracked by control chart methods³ subsequent to each sampling episode (i.e., 12 times per year). If concentrations are found to be "out of control," the data will be inspected more closely and, if necessary, adjustments will be made to the treatment system to account for increased concentrations of the compound.

Actions To Be Taken Based On Laboratory Results:

Concentrations of all key compounds will be tracked by control chart methods as described above.

Reporting:

Results will be presented in the first semi-annual report which is written subsequent to sampling as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) Control charts constructed from concentrations of key compounds will be presented as shown in Figure D-1.
- (4) Any adjustments made to the treatment system resulting from the presence of "out of control" compounds will be discussed.
- (5) If any compounds are added to the list of key compounds, the data and rationale which support the addition will be supplied.

³ Details of control chart methods which will be used to track concentrations for process control are presented in Appendix E-1.

INFORMATION GOAL #3: Determine the lateral extent of contamination plumes.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type: Wells near plume perimeters.

Number Of Wells To Be Sampled: 87

Sampling Frequency: 2, 4 or 12 per year.

Data Analysis Procedures:

The following procedures will be conducted on a yearly basis:

- (1) Annual medians of key compounds (Table C-4) (or sums of key compounds) will be plotted on a site map.
- (2) An interval estimate of the difference between medians for the current two years of data and the previous two years of data will be made for each key compound (or for each sum of key compounds). The estimates will indicate whether or not concentrations have changed enough in the well to warrant relocation of the plume boundary. Details of this statistical procedure are presented in Appendix E-2.
- (3) Plume boundaries will be drawn based on personal judgment and the above information. Knowledge of flow direction, site hydrogeology, and contaminant transport behavior will support "personal judgment" decisions.

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in each annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) Annual medians of key compounds (or sums of key compounds) will be plotted on a site map.
- (4) Interval estimates of the differences in medians between the current two years data and the previous two years data will be given for each key compound (or for each sum of key compounds) (see Figure D-2).
- (5) Plume boundaries for all six Areas of Concern (AOCs) will be drawn on a site map. Plume boundaries are defined as an annual median concentration equal to $5\mu/L$ of one key compound or sum of key compounds.

INFORMATION GOAL #4: Determine the mass of contaminants removed at each remediation area.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Extraction wells

Number Of Wells To Be Sampled: 11

Sampling Frequency: 12 per year

Data Analysis Procedures:

The volume of water pumped will be multiplied by the concentrations of each key compound (Table C-4) to obtain the mass of contaminants removed.

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in each annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) Masses of contaminants removed, as well as volume and concentration data used calculate the masses, will be reported.

INFORMATION GOAL #5: Characterize contaminant concentrations over time within the plumes.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Wells within the plumes.

Number Of Wells To Be Sampled: 11

Sampling Frequency: 2 or 12 per year.

Data Analysis Procedures:

Plots of time versus concentration will be constructed for all key compounds (Table C-4). The plots will be observed in order to improve current understanding of the relationship between remediation efforts and contaminant behavior.

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in each annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) Plots of time versus concentration for key compounds will be provided (see Figure D-3).

INFORMATION GOAL #6: Determine the hydraulic effectiveness of remediation.

Compounds To Be Analyzed: Not applicable because only water elevations will be measured.

Well Type To Be Sampled: All onsite (i.e., main site and west complex) monitoring wells.

Number Of Wells To Be Sampled: 149

Sampling Frequency: 4 per year

Data Analysis Procedures:

Water elevations will be used to construct contour maps of the water table. If necessary, pumping will be adjusted based information provided by the contour maps.

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in each annual report as follows:

- (1) All water elevations measured in monitoring wells will be reported in a raw data format.
- (2) Contour plots of water table elevations will be provided.

INFORMATION GOAL #7: Measure water quality down-gradient of AOCs where flow may pass across a site boundary.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Wells which are both near the property boundary and down-gradient of an AOC.

Number Of Wells To Be Sampled: 6

Sampling Frequency: 4 or 12 per year.

Data Analysis Procedures:

On a yearly basis, concentrations will be compared to the Ground Water Protection Standard (GWPS) using interval estimation. Details of the estimation procedure are presented in Appendix E-3.

If data analysis procedures indicate that the GWPS has been exceeded, IBM will attempt to determine the reason for the change in ground water quality, as well as implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.

Actions To Be Taken Based On Laboratory Results:

If a compound is detected which is at least twice the level of the GWPS, the well will be resampled. If the concentration is still at or above twice the GWPS, then the concentration is confirmed, and:

- (1) The pertinent well, compound and concentration will be reported to NYSDEC within seven days of confirmation.
- (2) IBM will attempt to determine the reason for the change in ground water quality, as well as implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.

Reporting:

Results will be presented in each annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) If a resample is conducted and the compound of interest is no longer at or above twice the GWPS, then the original analysis will be labeled "invalid."
- (4) If a resample is conducted and the concentration is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.
- (5) Interval estimates of annual concentration medians will be given for all routinely monitored compounds (see Figure D-4).
- (6) Documentation will be provided regarding actions, decisions or observations made as the result of either: (a) the discovery of interval estimates of annual medians which exceed the GWPS, or (b) the discovery of concentrations above twice the GWPS.

INFORMATION GOAL #8: Confirm previous monitoring results which indicate that contaminants are not migrating offsite from onsite.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Clean wells located near the property boundary.

Number Of Wells To Be Sampled: 14

Sampling Frequency: 2 per year

Data Analysis Procedures:

Concentrations will be compared with the GWPS.

Actions To Be Taken Based On Laboratory Results:

If a compound is detected at or above the GWPS, the well will be resampled. If the concentration is still at or above the GWPS, then the concentration is confirmed, and:

- (1) The pertinent well, compound and concentration will be reported to NYSDEC within seven days of confirmation.
- (2) IBM will immediately attempt to determine the reason for the change in ground water quality, as well as implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.
- (3) If appropriate, the well may be redesignated as a well for meeting goal #7.

Reporting:

Results will be presented in each semi-annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) If a resample is conducted and the compound of interest is no longer at or above the GWPS, then the original analysis will be labeled "invalid."
- (3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.
- (4) Any actions, decisions or observations made as the result of discovery of concentrations above the GWPS will be reported.
- (5) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

INFORMATION GOAL #9: Confirm previous monitoring results which indicate that contaminants are not migrating onsite from offsite.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Clean wells located near the property boundary.

Number Of Wells To Be Sampled: 8

Sampling Frequency: 2 per year

Data Analysis Procedures:

Concentrations will be compared with the GWPS.

Actions To Be Taken Based On Laboratory Results:

If a compound is detected at or above the GWPS, the well will be resampled. If the concentration is still at or above the GWPS, then the concentration is confirmed, and:

- (1) The pertinent well, compound and concentration will be reported to NYSDEC within seven days of confirmation.
- (2) IBM will immediately attempt to determine the reason for the change in ground water quality, as well as implications of the change in terms of protection of human health and the environment.

Reporting:

Results will be presented in each semi-annual report as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) If a resample is conducted and the compound of interest is no longer at or above the GWPS, then the original analysis will be labeled "invalid."
- (3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.
- (4) Any actions, decisions or observations made as the result of discovery of concentrations above the GWPS will be reported.
- (5) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

INFORMATION GOAL #10: Ensure that all chemicals of concern are properly identified and monitored.

Compounds To Be Analyzed: "Appendix 33" compounds (Table C-1).

Well Type To Be Sampled: Wells within the plumes.

Number Of Wells To Be Sampled: 8

Sampling Frequency: 1 per year for 3 years.

Data Analysis Procedures:

Concentrations will be compared with the MDL.

Actions To Be Taken Based On Laboratory Results:

If an additional⁴ compound is detected at or above the MDL, the well will be resampled. If the concentration is still at or above the MDL, then the presence of the compound is confirmed, and:

- (1) The pertinent well, compound and concentration will be reported to NYSDEC within seven days of confirmation.
- (2) The well will be sampled for two quarters following the original confirmation. Laboratory results will be reported to NYSDEC.
- (3) A decision will be made regarding whether or not to place the additional compound on the list of routinely monitored compounds (i.e., Table C-2). The decision will be based on several factors including the concentration of the compound in relation to the GWPS, and the location and direction of ground water flow.

Reporting:

Results will be presented in the first semi-annual report which is written subsequent to sampling as follows:

- (1) Results of all laboratory analyses will be reported in a raw data format.
- (2) If a resample is conducted and the compound of interest is no longer at or above the MDL, then the original analysis will be labeled "invalid."
- (3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

⁴ "Additional" refers to compounds other than those which are routinely monitored (i.e. other than those compounds listed in Table C-2).

INFORMATION GOAL #11: Provide a qualitative basis for comparison with data from "Appendix 33" monitoring of contaminated wells.

Compounds To Be Analyzed: "Appendix 33" compounds (Table C-1).

Well Type To Be Sampled: Clean wells located away from AOCs.

Number Of Wells To Be Sampled: 2

Sampling Frequency: 1 per year for 3 years.

Data Analysis Procedures:

To be determined.

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in the first semi-annual report which is written subsequent to sampling as follows:

(1) Results of all laboratory analyses will be reported in a raw data format.

INFORMATION GOAL #12: Provide data for investigative studies associated with the NYS Part 373 Corrective Action Program.

Compounds To Be Analyzed: Routinely monitored compounds (Table C-2).

Well Type To Be Sampled: Wells in areas under investigation.

Number Of Wells To Be Sampled: 64

Sampling Frequency: 4 per year

Data Analysis Procedures:

To be determined.⁵

Actions To Be Taken Based On Laboratory Results:

None

Reporting:

Results will be presented in the first semi-annual report which is written subsequent to sampling as follows:

- (1) Results of all laboratory analyses will be reported as raw data.
- (2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.
- (3) A summary of project status will be provided.
- (4) Annual medians of key compound(s) will be plotted on a site map.

⁵ A separate data analysis protocol will be written for each Remedial Facilities Investigation (RFI) work plan.

Appendix B

Sampling Frequencies

Appendix B Sampling Frequencies

Table B-1. Sampling frequencies (per year)—area A remediation area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
702			2			4						
007			2			4						
998			2			4						
999			2			4						
720			2			4						
761			2			4						
99			2			4						
763			2			4						
204			2			4						
714						4				1		
Sump	1	12				4						
16				2	2	4						
103				2	2	4						
104				2	2	4						
744						4						4
745						4						4
15						4						4
706						4				1		
SW						4						12

Table B-2. Sampling frequencies (per year)—area B remediation area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
863	1	12	12	12	12	4						
834			4			4						
571			4			4						
064			4			4	4					
065			4			4	4					
863 GAC		12				4						

Table B-3. Sampling frequencies (per year)—area C remediation area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
563			2			4						
561			2			4						
559			2			4				1		
053			2			4						
054			2			4						
932			2			4						
944			2			4						
943			2			4						
069			2			4						

Table B-4. Sampling frequencies (per year)—area D remediation area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
032	1	12	12	12	12	4						
757			4			4						
067			4			4	4					
066			4			4	4					

Table B-5. Sampling frequencies (per year)—SEQ remediation area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
048			4			4						
552			4			4				1		
543			4			4						
545			4			4						
990			12			4	12					
987			12			4	12					
520			4			4						
502			4			4						
503			4			4						
859						4						4
860						4						4
861						4						4
862						4						4

Table B-6. Sampling frequencies (per year)—area C B/330 investigative area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
574						4						1
178						4						4
576						4						4
921						4						2
922						4						2
923						4						2
924						4						2
954						4						2
955						4						2
956						4						2
957						4						2
965						4						2
966						4						2
941						4						2
PVV-7	1	12	12	12	12	4						
572						4						1
181						4				1		4
756						4						4
062						4						4
063						4						4
061						4						4
059						4						4
060						4						4
143						4						2
145						4						2
146						4						2
929						4						2
930						4						2
931						4						2
953						4						2

Table B-7. Sampling frequencies (per year)—building 322 investigative area.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
735						4						4
740						4						4
766						4						4
767						4						4
768						4						4
769						4						4
770						4						4
771						4						4
772						4						4
773						4						4
754						4						4
755						4						4
792						4						4
793						4						4
781						4						4
203						4						4
151						4						4
949						4						4
743						4						4
161						4						4
35						4						4
68						4						4
774						4						4
775						4						4
791						4						4
789						4						4
202						4						4

Table B-8. Sampling frequencies (per year)—bedrock aquifer.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
PW-1		12	12	12	12	4						
PW-2	1	12	12	12	12	4						
PW-4	1	12	12	12	12	4						
PW-5/5A		12	12	12	12	4						
PW-6		12	12			4						
PW-7	1	12	12	12	12	4						
PW-9 (when on)		12	12	12	12	4						
GAC-Input	1	12				4						
113			4			4						
535			4			4						
958			4			4						
515			4			4						
963			4			4						
964			4			4						
521			4			4						
857			4			4						
858			4			4						
166			2			4			2			
167			2			4			2			
087			2			4			2			
085			2			4			2			
722			2			4						
711			2			4						
105			2			4						
106			2			4						
704			2			4						
108			2			4						
109			2			4						
959			2			4				1		

Table B-8. Sampling frequencies (per year)—bedrock aquifer, continued.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
965			2			4						
966			2			4						
564			2			4						
563			2			4						
559			2			4						
932			2			4						
944			2			4						
943			2			4						
145			2			4						
146			2			4						
929			2			4						
930			2			4						
931			2			4						
941			2			4				1		
953			2			4						
143			2			4						
156			2			4						
945			2			4						
952			2			4						
737			2			4						
742			2			4						
173			2			4						
150			2			4		2				
779			2			4		2				
778			2			4		2				
777			2			4		2				
739			2			4						
716			2			4				1		

Table B-9. Sampling frequencies (per year)—west complex.

Well	Information Goal											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
407											1	
408											1	

Note: Monitoring wells for information goal #8 were selected after the initial submittal of the ground-water monitoring proposal and are therefore not included in the Appendix B tables.

Appendix C

Categories of Compounds

Table C-1. Appendix 33 compounds.

Appendix 33, found in 6 NYCRR Subpart 373-2¹, is a list of 17 inorganic compounds and 202 organic compounds. Various sections of the Subpart 373-2 regulations require that ground water samples be analyzed for the presence of Appendix 33 compounds. The regulations also describe actions which should be taken if Appendix 33 compounds are confirmed to be present.

Appendix 33 provides the compound's chemical name, the Chemical Abstracts Service Registry Number, the Chemical Abstracts Service Index Name, suggested analytical methods, and Practical Quantitation Limits (PQL's). Regulatory requirements pertain only to the list of substances. The analytical methods and PQL's are given solely for information purposes.

¹ New York State Department of Environmental Conservation, *Final Status Standards for Owners and Operators of Hazardous Waste Treatment, Storage and Disposal Facilities*, Albany, NY, January 31, 1992.

Table C-2. Routinely monitored compounds.

Constituents	CAS #
Acetone	67-64-1
Benzene ²	71-43-2
Bromochloromethane	74-97-5
Chlorobenzene ²	108-90-7
Chloroethane	75-00-3
Chloroform	67-66-3
Chloromethane	74-87-3
Dibromochloromethane	124-48-1
1,2-Dibromoethane	106-93-4
1,2-Dichlorobenzene ²	95-50-1
1,3-Dichlorobenzene ²	541-73-1
1,4-Dichlorobenzene ²	106-46-7
Dichlorodifluoromethane (Freon 12) ²	75-71-8
1,1-Dichloroethene ²	75-35-4
cis-1,2-Dichloroethene ²	156-59-2
Trans-1,2-Dichloroethene	156-60-5
1,3-Dichloropropane	142-28-9

Table C-2. Routinely monitored compounds, continued.

Constituents	CAS #
Ethylbenzene ²	100-41-4
Trichlorotrifluoroethane (Freon 113 or Freon TF) ²	76-13-1
Hexachlorobutadiene	87-68-3
Isopropylbenzene	98-82-8
p-Isopropyltoluene	99-878-6
Methylene Chloride ²	75-09-2
Methyl-t-Butyl Ether	1634-04-4
Naphthalene	106-44-5
n-Propylbenzene	103-65-1
1,1,2,2-Tetrachloroethene ²	127-18-4
THF	109-99-9
Toluene	108-88-3
1,2,3-Trichlorobenzene ²	95-63-6
1,2,4-Trichlorobenzene ²	96-18-4
1,1,2-Trichloroethene	79-01-6
Trichlorofluoromethane (Freon 11)	75-69-4
1,2,4-Trimethylbenzene	95-63-6
1,3,5-Trimethylbenzene	108-67-8
Vinyl Chloride ²	75-01-4
m-Xylene	108-38-3
p-Xylene	106-42-3
o-Xylene	95-47-6
1,1,1-Trichloroethane	71-55-6
n-Butylbenzene	104-51-8
sec-Butylbenzene	135-98-8
1,1-Dichloroethane	75-34-3
1,2-Dichloroethane	107-06-2
1,2-Dichloro-1,2,2-Trifluoroethane (Freon 123a) ²	354-23-4
Chromium (total) ²	7440-47-3
Fluoride ²	7782-41-4 ³
Zinc (total) ²	7440-66-6

² Listed in Table C-3.³ CAS # for Fluorine.

Table C-3. Selected compounds.

Constituents	Method ⁴	Conc. Limit (µg/L) ⁵
Benzene	8021	0.7
Chlorobenzene	8021	5.0
1,2-Dichlorobenzene ⁶	8021	4.7
1,2,3-Trichlorobenzene	8021	5.0
1,2,4-Trichlorobenzene	8021	5.0
1,3-Dichlorobenzene	8021	5.0
1,4-Dichlorobenzene ⁶	8021	4.7
Dichlorodifluoromethane (Freon 12)	8021	5.0
1,1-Dichloroethene	8021	5.0
Methylene Chloride	8021	5.0
Ethylbenzene	8021	5.0
Vinyl Chloride	8021	2.0
1,2-Dichloro-1,2,2-Trifluoroethane (Freon 123a)	8021	5.0
1,1,2,2-Tetrachloroethene	8021	5.0
Trichlorotrifluoroethane (Freon 113)	8021	5.0
1,1,2-Trichloroethene	8021	5.0
cis-1,2-Dichloroethene	8021	5.0
Fluoride ⁷	4500-F C ⁸	1500.0
Chromium (total) ⁹	7191	50.0
Zinc (total) ¹⁰	6010	300.0

⁴ EPA Method SW-846. Method 8010 or 8020 may be substituted for method 8021.

⁵ Ground water protection concentrations for corrective measures.

⁶ The sum of 1,2-Dichlorobenzene and 1,4-Dichlorobenzene must be ≤ 4.7 µg/L.

⁷ Area "A" and Area "C".

⁸ APHA, AWWA, WPCF, American Public Health Association. 1989. Standard Methods for the Examination of Water and Wastewater. 17th ed., Washington, DC, p. 4-87.

⁹ Area "A".

¹⁰ Area "C".

Table C-4. Key compounds.

Location	Compound
Area A remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Area B remediation area	Tetrachloroethene
Area C remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene Vinyl Chloride Freon TF
Area D remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
SEQ remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Area C B/330 investigative area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene Vinyl Chloride Freon TF
Building 322 investigative area	Freon TF Freon 123a Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Bedrock aquifer	Freon TF Freon 123a Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene

Appendix D

Reporting Formats

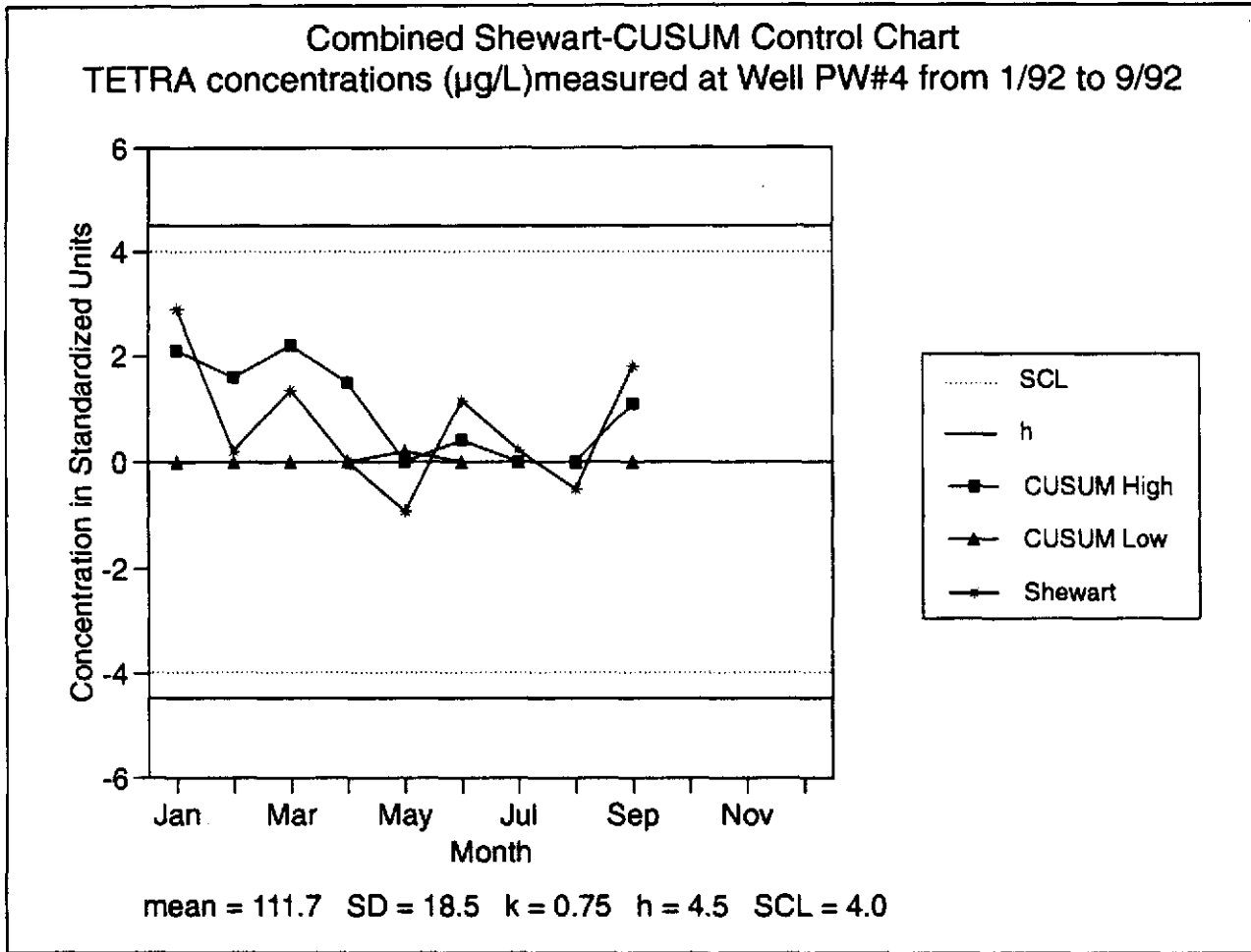


Figure D-1. Reporting format for combined Shewart-CUSUM control chart.

Well #	Compound	Group 1		Group 2		Δ ($\mu\text{g/l}$)	CI% ($\mu\text{g/l}$)	CL _l ($\mu\text{g/l}$)	CL _u ($\mu\text{g/l}$)
		dates	n	dates	m				
016	CEDC	88-89	4	90-91	4	0.95	89	-0.5	2.4

Figure D-2. Reporting format for Hodges-Lehmann estimates and confidence intervals.

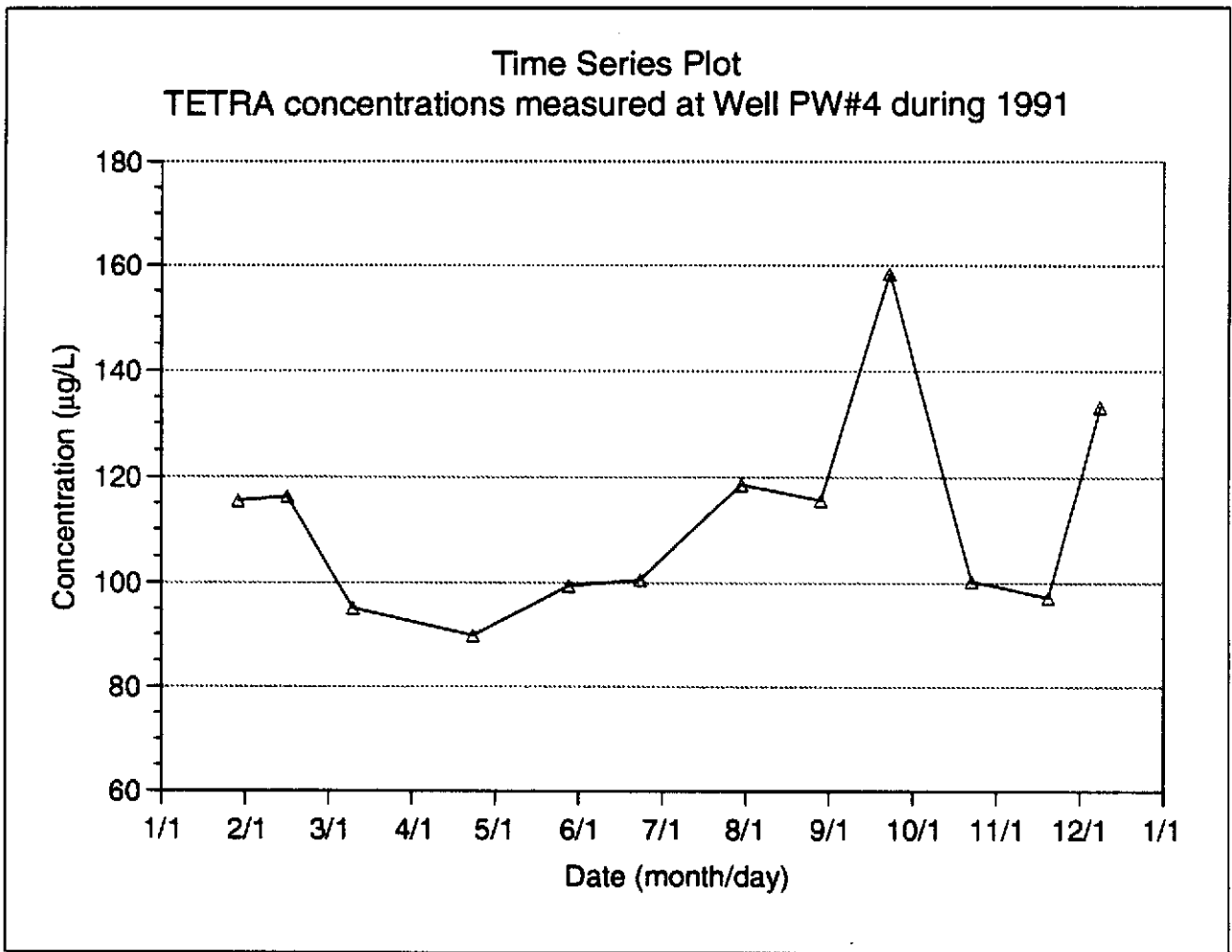


Figure D-3. Reporting format for time series plots.

Well #	Compound	Date	n	CIX	CL ₁	CL ₂
PW#4	TRICE	1986	12	96	2.1	2.4

Figure D-4. Reporting format for interval estimates of the median.

Appendix E

Data Analysis Methods

INTRODUCTION

The design and implementation of combined Shewart-CUSUM control charts is described in this appendix. Definitions of terms and parameters are presented first, followed by design procedures. The procedures are then used to design the control chart scheme for information goal #2: *ensure adequacy of the treatment facility by detecting changes in input concentrations*. Implementation procedures are outlined and then applied to data obtained from the IBM East Fishkill facility. Finally, a table of references for control charts is given.

It is assumed that each contaminant at each well will be tracked separately. It is also assumed that, for the most part, only one observation will be obtained per sampling event. Equations are provided, however, which accommodate average values obtained from multiple observations obtained for quality control purposes.

DEFINITIONS OF TERMS

ARL (average run length): The average number of samples before an out-of-control signal is obtained. The ARL should be large when the process is close to the target value and small when the process has shifted too far from the target value.

FIR (fast initial response) feature: Assigning a value greater than 0 to the CUSUM "headstart value", S_0 .

out-of-control: The situation which exists when the process has shifted too far from the target value.

two-in-a-row rule: A process is declared out-of-control only if two out-of-control signals are obtained in succession.

DEFINITIONS OF PARAMETERS

- μ Estimate of the population mean. It is used as the target value for the Shewart-CUSUM quality control scheme.
- σ Estimate of the population standard deviation.
- SCL Shewart control limit.
- k A parameter of the CUSUM scheme known as the "reference value."
- h A parameter of the CUSUM scheme known as the "decision interval."
- Δ The deviation to be detected by the CUSUM scheme.
- S_{Hi} The CUSUM statistic at time = i for positive shifts.
- S_{Li} The CUSUM statistic at time = i for negative shifts.
- S_0 The CUSUM statistic at time = 0. Also known as the "headstart value."

- Y_i The i th observation for both the Shewart and CUSUM schemes. It represents a single reading.
- \bar{Y}_i The i th observation for both the Shewart and CUSUM schemes. It represents the average from a number of observations.
- Z_i The standardized value of the i th observation (or average of observations). It is plotted on the control chart and compared to the SCL . It is also used in calculating S_{Hi} and S_{Li} .

ommended for situations where outliers are likely. It is particularly suitable if only single observations rather than an average of observations are used, because the effects of an outlier will be smoothed out if averages are taken.

The two outliers do not have to be from the same side of the distribution. Although a suspected outlier on the high side followed by a suspected outlier on the low side may not indicate a shift in the process mean, it would indicate the need for closer examination of water quality data.

DESIGN PROCEDURES

1. Use the worksheet shown in Figure E-1 to design the control chart.
2. State the monitoring information goal, monitoring approach, and statistical information goal.
3. **Decide whether to use one-sided or two-sided charts.** Two-sided control charts should be used if both increases and decreases in concentration are of interest.
4. Choose a value for SCL .
5. **Choose a value for k .** For a given in-control ARL, the quickest detection is obtained when $k = \Delta/2$.
6. **Choose a value for h .** h can be chosen from a table of average run lengths (ARLs). The value of h should be selected to give the desired ARL when the process is in-control and when it is out-of-control.
7. **Choose a value for S_o .** If the FIR feature is not used, $S_o = 0$. If the FIR feature is used, S_o is frequently set at $h/2$. The FIR feature should be implemented if it is felt that the process is likely to be out-of-control at startup or after a restart following a control action.
8. **Decide whether or not to use the two-in-a-row rule.** The two-in-a-row rule is rec-

APPLICATION OF DESIGN PROCEDURES

The above procedures were used to design the control chart for information goal #2. The design is summarized in Figure E-2.

IMPLEMENTATION PROCEDURES

1. Use the worksheet in Figure E-3 to implement the control chart.
2. **Calculate μ and σ from historical data.** Use data which was collected when the system was operating within desirable limits. Outliers, as well as data which is part of an obvious trend, should not be used. Precision of the estimates will increase with increasing sample size.
3. **Construct the control chart.** The horizontal axis should be in units of time and the vertical axis in standardized concentration units. Draw in the upper Shewart limit at SCL , lower Shewart limit at $-SCL$, upper CUSUM limit at h , and lower CUSUM limit at $-h$.
4. Calculate:

$$Z_i = (Y_i - \mu) / \sigma \text{ for a single observation}$$

$$Z_i = (\bar{Y}_i - \mu) / \sigma \text{ for multiple observations}$$

5. Calculate:

$$S_{Hi} = \max [0, (Z_i - k) + S_{H(i-1)}]$$

$$S_{Li} = \max [0, (-Z_i - k) + S_{L(i-1)}]$$

6. Plot Z_i , S_{Hi} and $-S_{Li}$ on the control chart.

7. An out-of-control signal is given if $|Z_i| \geq SCL$ or if S_{Hi} or $S_{Li} \geq h$.

8. Declare the process out-of-control if two out-of-control signals occur in succession.

9. If any parameters of the control chart are modified, S_{Hi} and S_{Li} should be reset to 0.

APPLICATION OF IMPLEMENTATION PROCEDURES

The above procedures were used to implement the control chart which was designed for information goal #2. Implementation is summarized in Figure E-4.

UPDATING AND REVISING PROCEDURES

There are three situations where the Shewart-CUSUM control chart may need to be updated or revised.

1. **Recalculate μ and σ after the process is declared out-of-control.** Control charts were traditionally used for industrial or manufacturing purposes. If a control chart indicated that a process was out-of-control, adjustments were made to the process itself.

The scenario is somewhat different for the information goal stated here because the "process" cannot be modified. Input concentrations are monitored to ensure adequacy of the treatment system. If the control chart indicates that the process is out-of-control, the treatment system or perhaps the flow are adjusted, not the input concentrations.

For example, if the control chart indicates a persistent increase in TETRA concentrations, the treatment system may have to be modified to handle the larger concentrations. At that point it might be appropriate to revise μ (and perhaps σ) prior to restarting the control chart.

For the information goal listed here, μ and σ will be examined and, if necessary, recalculated after each out-of-control episode.

2. **Improve the precision of μ and σ as more data becomes available.** New data should periodically be combined with the original data which was used to calculate μ and σ . The parameters can then be updated. For the information goal listed here, μ and σ will be updated annually.

3. **Revise control chart parameters to better serve information needs.** There are many ways to revise Shewart-CUSUM control charts to meet information needs. For example, if the control chart is too sensitive, values of SCL , h and/or k can be increased. This type of revision will be conducted according to section VII of the data analysis protocol.

Worksheet for Designing Shewart-CUSUM Control Charts

Monitoring Information Goal:

Monitoring Approach:

Statistical Information Goal:

One-sided or Two-sided?

Parameter	Value	How Obtained

Decision Rule:

Comments:

Figure E-1. Worksheet for designing combined Shewart-CUSUM control charts.

Worksheet for Designing Shewart-CUSUM Control Charts

Monitoring Information Goal: Ensure adequacy of the treatment facility by detecting increases in input concentrations.

Monitoring Approach: Detect increases in mean concentration over time which are moderate and persistent or large and sudden.

Statistical Information Goal: Immediately detect concentrations which are $\geq 4\sigma$ above the target mean. Quickly detect persistent changes in concentration which are $\geq 1.5\sigma$ above the target mean.

One-sided or Two-sided? two-sided

Parameter	Value	How Obtained
SCL	4.0	information goal
k	0.75	$k = \Delta/2$: $\Delta = 1.5$ as per information goal
h	4.5	Part 1, Table 2 in Lucas (1982)
S_o	0.0	process not likely to be out-of-control at startup

Decision Rule: two-in-a-row rule

Comments: The two-in-a-row rule was chosen because outliers are likely.

Figure E-2. Design of combined Shewart-CUSUM control chart for information goal #2.

Worksheet for Implementing Shewart-CUSUM Control Charts

Well #: _____ Compound: _____

$\mu =$ _____ $\sigma =$ _____

Concentrations ($\mu\text{g/L}$) used for calculating μ and σ					
Date	Conc.	Date	Conc.	Date	Conc.

Month	Y_i or \bar{Y}_i ($\mu\text{g/L}$)	Z_i	$Z_i - k$	S_{Hi}	$-Z_i - k$	S_{Li}
Jan						
Feb						
Mar						
Apr						
May						
Jun						
Jul						
Aug						
Sep						
Oct						
Nov						
Dec						

Comments:

Figure E-3. Worksheet for implementing combined Shewart-CUSUM control charts.

Worksheet for Implementing Shewart-CUSUM Control Charts

Well #: PW#4

Compound: Tetrachloroethene (TETRA)

 $\mu = 111.7 \text{ } \mu\text{g/L}$ $\sigma = 18.5 \text{ } \mu\text{g/L}$ Concentrations ($\mu\text{g/L}$) used for calculating μ and σ

Date	Conc.	Date	Conc.	Date	Conc.
01-28-91	115.5	06-24-91	100.5	11-20-91	97.2
02-15-91	116.3	07-31-91	118.6	12-09-91	133.2
03-11-91	95.1	08-29-91	115.6		
04-24-91	89.8	09-23-91	158.4		
05-29-91	99.4	10-23-91	100.3		

Month	Y_i or \bar{Y}_i ($\mu\text{g/L}$)	Z_i	$Z_i - k$	S_{Hi}	$-Z_i - k$	S_{Li}
Jan	165.0	2.88	2.13	2.1	-3.63	0
Feb	116.4	0.25	-0.50	1.6	-1.00	0
Mar	137.0	1.36	0.61	2.2	-2.11	0
Apr	112.0	0.02	-0.73	1.5	-0.77	0
May	94.6	-0.92	-1.67	0	0.17	0.2
Jun	133.1	1.16	0.41	0.4	-1.91	0
Jul	115.8	0.22	-0.53	0	-0.97	0
Aug	102.5	-0.50	-1.25	0	-0.25	0
Sep	145.4	1.82	1.07	1.1	-2.57	0
Oct						
Nov						
Dec						

Comments: Z_i , S_{Hi} and $-S_{Li}$ are plotted on the control chart shown in Figure E-5. No out-of-control signals occurred.

Figure E-4. Implementation of combined Shewart-CUSUM control chart for information goal #2.

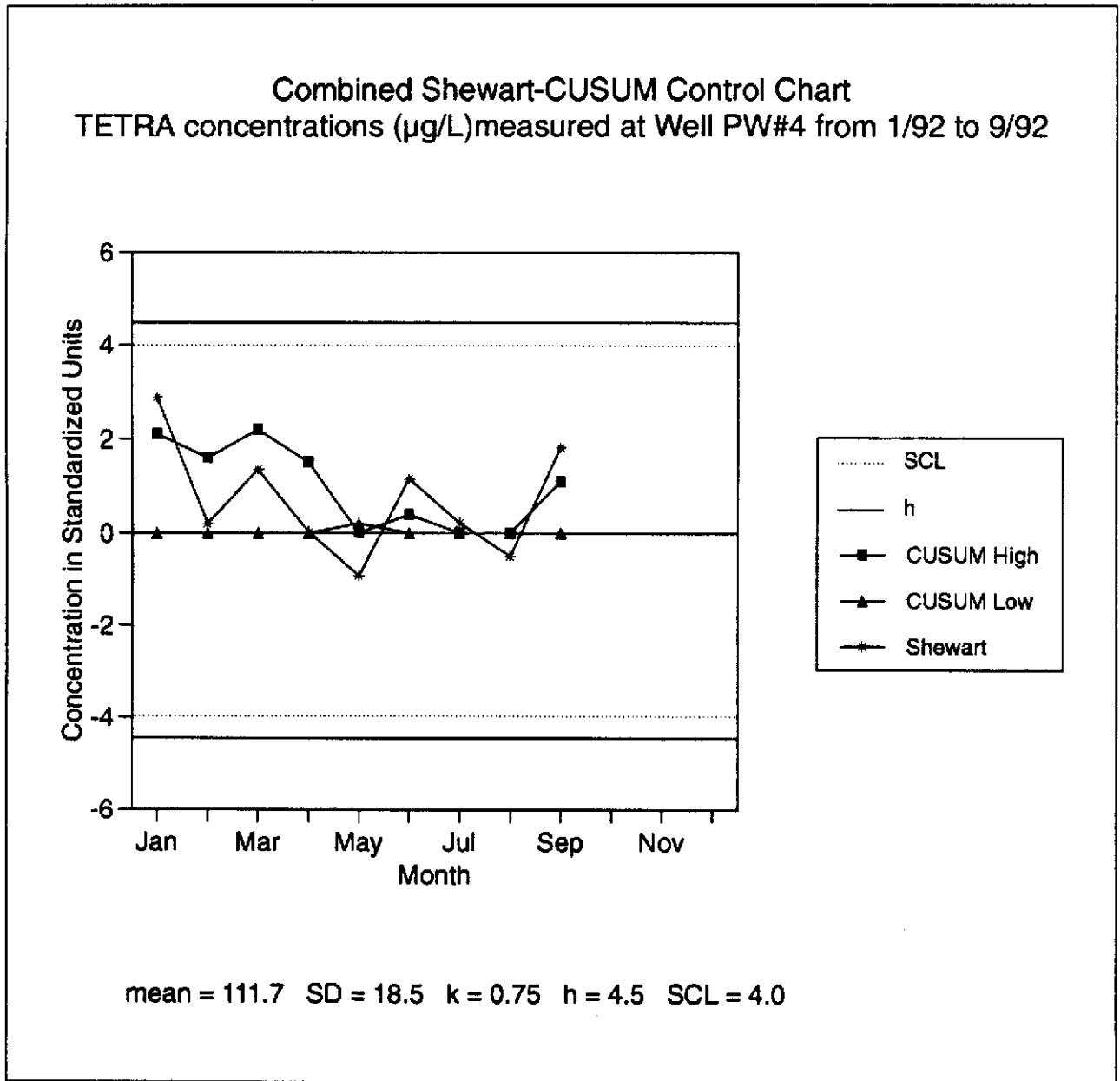


Figure E-5. Plot of Shewart-CUSUM data from Figure E-4.

REFERENCES

Reference	Comments
Berthouex and Hunter, 1975	Discusses the use of Shewart charts for monitoring sewage treatment plants.
Berthouex et al., 1978	Examines the effect of data record attributes on the use of Shewart and CUSUM charts for monitoring sewage treatment plants.
Bissell, 1984	Evaluates the performance of control charts in the presence of serial correlation.
Cheremisinoff, 1988	Describes how to construct control charts using Lotus 1-2-3®.
Hockman and Lucas, 1987	A well-written explanation of CUSUM schemes. Also describes how to use MINITAB® to construct CUSUM charts.
Lucas, 1982	An excellent article on combined Shewart-CUSUM control charts.
Lucas, 1985	Discusses the FIR feature and the two-in-a-row rule.
Lucas and Crosier, 1982	Discusses the two-in-a-row rule.
Page, 1954 and 1961	Describes the development of the CUSUM scheme.
Schilling and Nelson, 1976	Examines the effect of nonnormality on the control limits of Shewart charts.
Starks, 1989	Evaluates the use of control chart methods for RCRA sites.
Starks and Flatman, 1991	Basically the same as Starks, 1989.
U.S. EPA, 1989	Section 7 is devoted to the use of control charts at RCRA facilities.
Vasilopoulos and Stamboulis, 1978	Describes how to modify control chart limits in the presence of data correlation.
Vaughan and Russell, 1983	Discusses the use of control charts to monitor point source discharge.

Appendix E-2

The Hodges-Lehmann Estimator and Confidence Interval

INTRODUCTION

The Hodges-Lehmann estimator, $\hat{\Delta}$, is a nonparametric estimate of the difference between two independent groups. $\hat{\Delta}$ is the median of all possible pairwise differences between the x and y values.

DEFINITIONS OF PARAMETERS

- $\hat{\Delta}$ Hodges-Lehmann estimator:
 $\hat{\Delta} = \text{median} [x_i - y_i]$
- x_i The i th observation of group 1.
- y_i The i th observation of group 2.
- n The number of observations in group 1.
- m The number of observations in group 2.
- N The number of pairwise differences:
 $N = n \times m$.
- x^* The critical value from a table for the rank sum test. Used to calculate R_l and R_u for small sample sizes.
- $Z_{\alpha/2}$ The critical value from a table of standard normal quantiles. Used to calculate R_l and R_u for large sample sizes.
- R_l The rank of the pairwise difference which is the lower confidence limit of $\hat{\Delta}$.
- R_u The rank of the pairwise difference which is the upper confidence limit of $\hat{\Delta}$.

CL_l The lower confidence limit of $\hat{\Delta}$.

CL_u The upper confidence limit of $\hat{\Delta}$.

INFORMATION GOALS

Monitoring Information Goal: Determine the lateral extent of contamination plumes.

Statistical Approach: Compare recent concentration measurements to older measurements to see how much change there has been in contaminant concentration.

Statistical Information Goal: Estimate the magnitude of the difference between the current two years of concentration data and the previous two years of concentration data. Determine a 90 percent confidence interval for the estimate.

CALCULATION PROCEDURES

1. Calculate all pairwise differences.
2. Rank the pairwise differences from smallest to largest.
3. Calculate the median of the pairwise differences. This is the Hodges-Lehmann estimator.
4. If $m + n \leq 20$, go to step 5. Otherwise, go to step 7.
5. Refer to a table of quantiles for the rank sum statistic. Find the critical value x^* nearest to $\alpha/2$.

Appendix E-2 The Hodges-Lehmann Estimator and Confidence Interval

6. Calculate:

$$R_l = x^* - \frac{n(n+1)}{2}$$

$$R_u = N - R_l + 1$$

8. Calculate:

$$R_l = \frac{N - Z_{\alpha/2} \sqrt{\frac{N(n+m+1)}{3}}}{2}$$

$$R_u = N - R_l + 1$$

7. Refer to a table of standard normal quantiles. Find $Z_{\alpha/2}$.

9. Find the pairwise differences which correspond to R_l and R_u . These values are CL_l and CL_u , the upper and lower bounds of the confidence interval. The true difference will lie between CL_l and CL_u an average of $(100-\alpha)$ percent of the time.

EXAMPLE

Well #: 016

Compound: cis-1,2-Dichloroethene (CEDC)

Group 1		Group 2	
date	conc. (µg/L)	date	conc. (µg/L)
03-03-88	7.9	03-01-90	5.4
09-08-88	6.4	09-06-90	5.8
03-08-89	6.8	03-07-91	5.5
09-14-89	5.5	09-09-91	6.9

Pairwise Differences		Ordered Pairwise Differences	
2.5	1.4	-1.4	1.0
2.1	1.0	-0.5	1.0
2.4	1.3	-0.3	1.0
1.0	-0.1	-0.1	1.3
1.0	0.1	0.0	1.4
0.6	-0.3	0.1	2.1
0.9	0.0	0.6	2.4
-0.5	-1.4	0.9	2.5

median: $(0.9 + 1.0)/2 = 0.95 \text{ } \mu\text{g/L}$

$x^* = 12$ (for $\alpha = 0.114$)

$R_1 = 2$

$R_u = 15$

$CL_1 = -0.5 \text{ } \mu\text{g/L}$

$CL_u = 2.4 \text{ } \mu\text{g/L}$

$\hat{\Delta} = 0.95$ with an 89 percent confidence interval of -0.5 to 2.4.

REFERENCES

Reference	Comments
Helsel and Hirsch, 1992	Explains the Hodges-Lehmann estimator and confidence interval. Includes tables (Table B4).
Hirsch, 1988	Examines robustness of the Hodges-Lehmann estimator. Develops the seasonal Hodges-Lehmann estimator.
Hodges and Lehmann, 1963	Presents the original development of the Hodges-Lehmann estimator.
Hollander and Wolfe, 1973	Explains the Hodges-Lehmann estimator and confidence interval. Includes tables (Table A5).

INTRODUCTION

A nonparametric interval estimate of the true population median is described here. The estimate is calculated using the binomial distribution.

DEFINITIONS OF PARAMETERS

- x_i The i th observation of the sample.
- n The number of observations in the sample.
- x' The critical value from a table for the sign test (or a binomial table). Used to calculate R_l and R_u for small sample sizes.
- $Z_{\alpha/2}$ The critical value from a table of standard normal quantiles. Used to calculate R_l and R_u for large sample sizes.
- R_l The rank of the observation which is the lower confidence limit of $C_{0.5}$.
- R_u The rank of the observation which is the upper confidence limit of $C_{0.5}$.
- $C_{0.5}$ The true population median.
- CL_l The lower confidence limit of $C_{0.5}$.
- CL_u The upper confidence limit of $C_{0.5}$.

INFORMATION GOALS

Monitoring Information Goal: Measure water quality down-gradient of AOCs where flow may pass across a site boundary.

Statistical Approach: Determine the range of concentrations which will contain the true population median a large percent of the time.

Statistical Information Goal: On an annual basis, estimate the 95 percent confidence interval of the true population median concentration.

CALCULATION PROCEDURES

- Rank the sample observations from smallest to largest.
- If $n \leq 20$, go to step 3. Otherwise go to step 5.
- Refer to a table of quantiles for the sign test statistic. (If a binomial table is used, enter it at the $p = 0.5$ column). find the critical value x' nearest to $\alpha/2$.
- Calculate:

$$R_l = x' + 1$$

$$R_u = n - x' = x$$
- Refer to a table of standard normal quantiles. Find $Z_{\alpha/2}$.

6. Calculate:

$$R_l = \frac{n - Z_{\alpha/2} \sqrt{n}}{2}$$

$$R_u = \frac{n + Z_{\alpha/2} \sqrt{n}}{2} + 1$$

7. Find the observations which correspond to R_l and R_u . These values are CL_l and CL_u , the upper and lower bounds of the confidence interval. The true population median will lie between CL_l and CL_u an average of $(100-\alpha)$ percent of the time.
8. Declare the concentrations out of compliance if the entire confidence interval lies above the GWPS (see Figure E-6).

EXAMPLE

Well #: PW#4

Compound: Trichloroethene (TRICE)

Observations		Observations	
date	conc. (µg/L)	date	conc. (µg/L)
01-13-86	2.4	07-02-86	2.4
02-19-86	2.6	08-06-86	2.2
03-11-86	2.1	09-25-86	2.6
04-15-86	2.0	10-27-86	2.4
05-14-86	2.3	11-24-86	2.4
06-19-86	2.2	12-16-86	2.1

Ordered Observations	
2.0	2.4
2.1	2.4
2.1	2.4
2.2	2.4
2.2	2.6
2.3	2.6

$$x' = 2 \quad (\text{for } \alpha = 0.0193)$$

$$R_l = 3 \qquad R_u = 10$$

$$CL_l = 2.1 \text{ µg/L} \qquad CL_u = 2.4 \text{ µg/L}$$

The 96 percent confidence interval of the true population median is 2.1 to 2.4.

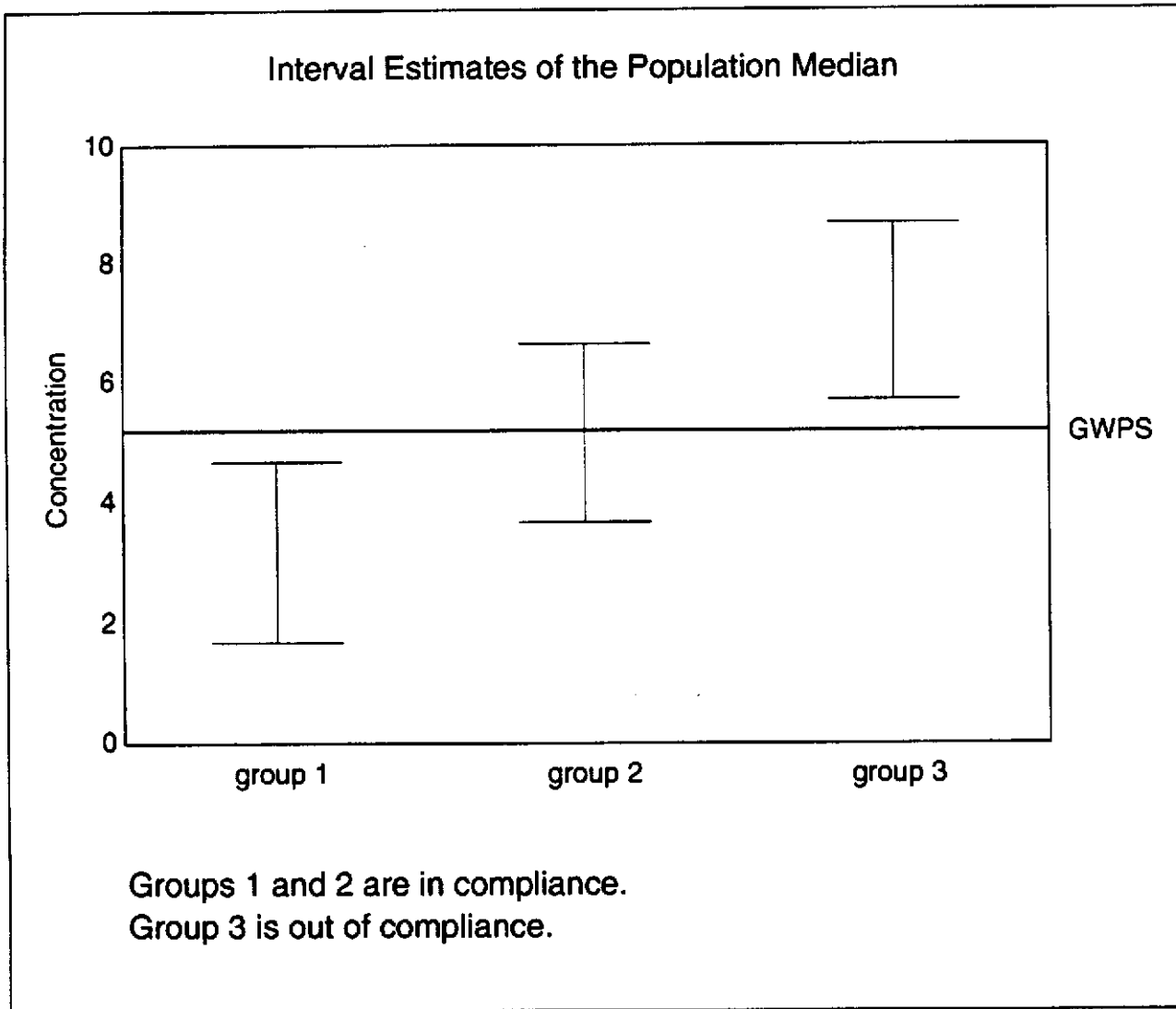


Figure E-6. Comparison of confidence intervals to the ground water protection standard (GWPS).

REFERENCES

Reference	Comments
Helsel and Hirsch, 1992	Explains how to obtain the nonparametric interval estimate of the median. Includes tables (Table B5).